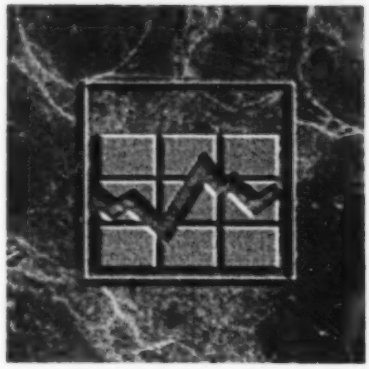
Catalogue no. 62F0026MIE — No. 001
ISSN: 1708-8879
ISBN: 0-662-35536-9

Research Paper

Household expenditures research paper series

# Survey of Household Spending 2001 - Data Quality Indicators

2001



Household Survey Methods Division
R.H. Coats Building, Ottawa, K1A 0T6
Telephone: 613 951-7355

Statistics Canada Statistique Canada

Canada

## How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to Client Services, Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

| | |
|---|---|
| **National inquiries line** | **1 800 263-1136** |
| **National telecommunications device for the hearing impaired** | **1 800 363-7629** |
| **Depository Services Program inquiries** | **1 800 700-1033** |
| **Fax line for Depository Services Program** | **1 800 889-9734** |
| **E-mail inquiries** | **infostats@statcan.ca** |
| **Web site** | **www.statcan.ca** |

## Ordering and subscription information

This product, Catalogue no. 62F0026MIE2003001, is available on Internet free. Users can obtain single issues at: http://www.statcan.ca/cgi-bin/downpub/research.cgi.

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Statistics Canada
Income Statistics Division

**Household expenditures research paper series**

# Survey of household spending 2001 - Data quality indicators

2001

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2003

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

November 2003

Catalogue no. 62F0026MIE2003001

ISSN: 1708-8879

ISBN: 0-662-35536-9

Frequency: Irregular

Ottawa

La version française de cette publication est disponible sur demande (n° 62F0026MIF au catalogue).

**Note of appreciation**

*Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.*

This report was prepared by the members of the Survey of Household Spending unit of the Household Survey Methods Division (HSMD):

Sylvie DeBlois
José Gaudet
Bruno Lapierre
Jessica Law
Christian Nadeau
Sylvain Nadon

The authors wish to thank the team responsible for the Survey of Household Spending in the Income Statistics Division (ISD) for their co-operation in the preparation of this report.

# Table of contents


**Highlights** ..... 5

**Introduction** ..... 7

**1. Sampling errors** ..... 8
- 1.1 Measures of Sampling Error ..... 8
- 1.2 Coefficients of Variation ..... 8
  - Table 1.1 ..... 9
  - Coefficients of Variation (%) by Province, Territory and at the National Level for the Estimation of Average Household Expenditures for Several Summary Level Expenditure Categories and for the Estimation of Average Income ..... 9
  - Table 1.2 ..... 10
  - Coefficients of Variation (%) at the National, Provincial and Territorial Level for Some Characteristics of Dwelling and Household Equipment ..... 10
- 1.3 Model for Deriving an Approximation of the CV ..... 10
- 1.4 Suppression of Unreliable Data in Estimation Tables ..... 12

**2. Nonresponse** ..... 12
- 2.1 Nonresponse Rates and Vacancy Rates ..... 12
  - Table 2.1 ..... 13
  - Nonresponse Rates (%) and Vacancy Rates (%) by Province, Territory and at the National Level ..... 13
- 2.2 Nonresponse According to Urbanization Level ..... 14
  - Table 2.2 ..... 14
  - Nonresponse Rates (%) and Vacancy Rates (%) by Urbanization Level ..... 14
- 2.3 Nonresponse According to Income Strata ..... 15
  - Table 2.3 ..... 15
  - Comparison of Nonresponse and Vacancy Rates (%) in High-income and Low-income Strata in Relation to Other Strata ..... 15
- 2.4 Adjustment for Nonresponse ..... 16

**3. Coverage errors** ..... 16
- 3.1 Undercoverage and Overcoverage: Slippage Rates ..... 17
  - Table 3.1 ..... 18
  - National Slippage Rates by Age-Sex Group ..... 18
  - Table 3.2 ..... 19
  - Slippage Rates for Provinces and Territories by Age-Sex Group ..... 19
  - Table 3.3 ..... 20
  - Slippage Rates for Provinces and Territories by Household Size ..... 20
- 3.2 Adjustment for Coverage Error at the Population Level and Household Level ..... 20

**4. Response errors** ..... 21

# Table of contents (continued)

**5. Processing errors** ..... **23**
5.1 Proportion of Households or Persons Requiring Imputation at National, Provincial and Territorial Level ..... 25
Table 5.1 ..... 26
Households Requiring Expenditure Imputation by Province and Territory ..... 26
Table 5.2 ..... 27
Persons Requiring Income Imputation and Persons Requiring Clothing Expenditure Imputation by Province and Territory ..... 27

**References** ..... **28**

**Appendix A Algebraic notation** ..... **29**

# Highlights

## Sampling Errors

- The coefficients of variation (CVs) of the average estimates of total expenditure per household vary between 1.2% and 1.9% for the provinces, and it is 0.7% at the national level. The CVs are higher in the territories, namely 5.8% in the Yukon, 5.7% in the Northwest Territories and 8.1% in Nunavut.

- The coefficients of variation (CVs) of the average estimates for the different summary level expenditure categories are less than or equal to 2.4% at the national level, and generally lower than 6% at the provincial level. The results for the characteristics of dwelling and household equipment are similar. The CVs at the territorial level are generally higher.

## Nonresponse

- The nonresponse rate is 23.8%. It is due to refusals (15.2%), to households that could not be contacted (5.2%), and finally to households with data that were considered unusable (3.4%).

- The final nonresponse rate generally increases with the urbanization level. The nonresponse rate is 19.8% at the rural level, and 30.6% in the urban centres of a million residents or more. This tendency is also observed in the collection nonresponse rates.

- Analysis of final response rates in the strata consisting of high and low-income geographic areas created under the sample design indicates that the nonresponse rate in high-income strata (37.5%) is higher than the rates observed in regular strata (23.5%) and in low-income strata (18.6%). The refusal rate for high-income strata is nearly twice the rate of low-income strata.

## Coverage Errors

- There is an undercoverage of 6% for households at the national level. Undercoverage of households is also observed for all provinces and territories, with rates varying from 4.1% to 21.6%. The Northwest Territories and Nunavut have the highest undercoverage of households.

- There is an undercoverage of 7.3% for persons at the national level. Undercoverage of persons is also observed for all provinces, with rates varying from 4.7% to 21.3%. The Northwest Territories and Nunavut have the highest undercoverage of persons.

- The national slippage rates for children (0 to 6 and 7 to 17) are quite different from those of other age groups. Overcoverage or a slight undercoverage occurs with respect to children, while there is always undercoverage among adults. The undercoverage rate for all children combined is 0.7%, while it is 9.2% for adults.

## Response Errors

- Response errors include, among others, recall errors, telescopic error and errors due to proxy response. Because the SHS interview is lengthy, the response burden can lead to respondent fatigue and have an impact on the data quality. Total interview time varies according to the household characteristics. For some households the interview can take more than five hours. The average length of interview was one hour and ten minutes.

## Processing Errors Related to Imputation

### i) Expenditure Variables

- 7.2% of respondents required some expenditure imputation (excluding the clothing section) with the majority of them having only one or two fields imputed out of the 251 expenditure variables.

- About 13% of the individuals required imputation for clothing variables. For the majority of these, the respondents provided the totals and only the components were imputed.

### ii) Income Variables

- Less than 4% of individuals required imputation for at least one income variable. For about 80% of these, total income was provided by the respondent and imputation was performed to obtain the breakdown by components.

# Introduction

The Survey of Household Spending (SHS) is an annual survey that collects data on household income and expenditure using personal interviews. The 2001 SHS sample is made up of 25,740 dwellings[1] distributed throughout the 10 provinces and the three territories. Collection takes place in January, February and March, and income and spending figures are obtained for the period from January 1 to December 31 of the previous year. Following a redesign that took place in 1997, this survey replaces the periodic Family Expenditure Survey and the Household Facilities and Equipment Survey (with modifications to questionnaires and samples).

Like all surveys, the SHS is subject to errors, despite all the precautions taken at the different stages of the survey to control them. While there is no comprehensive measure of the quality of the data generated by a survey, some quality measures produced at the different stages of the survey can provide users with the information needed in order to interpret the data properly.

This report therefore seeks to describe the quality indicators produced for the 2001 Survey of Household Spending. It covers the usual quality indicators that generally help users interpret data, such as coefficients of variation, nonresponse rates, slippage rates and imputation rates.

Quality indicators have been classified according to the main types of error encountered in a survey. Section 1 deals with sampling errors—that is, errors due to the fact that the inferences about the population drawn from the survey are based on information collected from a sample of the population, rather than the entire population. The following sections cover errors not due to sampling. Nonresponse and coverage errors are first discussed in sections 2 and 3. Response errors and processing errors are dealt with in sections 4 and 5 respectively.

This report focuses on data quality. For a detailed description of the methodology of the survey, see reference [1].

[1] From these 25,740 dwellings, we have to identify and exclude the ineligible dwellings (see section 2.1) to obtain the 22,172 households from which we collect data on household income and expenditure.

# 1. Sampling errors

Sampling errors exist when inferences about the population are drawn from the survey using information collected from a sample, rather than the entire population. In addition to the sample design and the estimation method used in the Survey of Household Spending, the sample size and the variability of each characteristic are factors that determine sampling error. Characteristics that are rare or are distributed very unevenly in the population will have greater sampling error than characteristics that are observed more frequently or are more homogeneous in the population.

## 1.1 Measures of Sampling Error

The standard error is a commonly used measure of sampling error. The standard error is the degree of variation of the estimate considering that a particular sample was selected, rather than another, among all possible samples of the same size under the same sample design. Since the SHS uses a complex sample design and estimation method, the standard error is estimated using a resampling method known as the Jackknife technique. For more details on this method, see reference [2].

The coefficient of variation (CV) is also a frequently used measure of the reliability of an estimate. It simply expresses the standard error as a percentage of the estimate. Thus, if an estimate Y is obtained for a certain characteristic and SE is the estimated standard error, then the CV will be (SE/Y) x 100.

Finally, either the standard error or the coefficient of variation may be used to derive another measure of the accuracy of estimates, namely the confidence interval. This measure indicates the level of confidence that, for a characteristic observed, the true value for the population lies within the interval. An interval with a confidence level of 95% corresponds to the estimate obtained from the sample ± 2 standard errors: (Y ± 2 SE).[2] This means that if the sampling were repeated a large number of times, each sample would provide a different interval and 95% of the intervals would contain the true value of the characteristic. Similarly, if the sampling were repeated, the interval Y ± SE would contain the true value in 68% of cases.

## 1.2 Coefficients of Variation

Estimates of coefficients of variation are calculated for estimates of many characteristics collected in the SHS. The CVs of detailed average household expenditure, as well as the CVs of dwelling characteristics and household facilities and equipment, are available at the national and provincial levels in the publication *User Guide—Survey of Household Spending* (see reference [3]).

It should be noted that the estimated CVs do not consider the fact that some of the data were imputed and thus may underestimate the true CVs. For most variables, the imputation rates are low (see section 5) and the provided CVs represent good estimates. To assess the reliability of detailed expenditure with a high imputation rate, the CV and the imputation rate should be considered simultaneously.

[2] The confidence interval is calculated directly from the CV in similar fashion, namely Y ± 2 (CV x Y)/100.

Table 1.1 gives an overview of the CVs of estimates of household averages for a few of the summary level expenditure categories and for income at the provincial, territorial and national level.

**Table 1.1**
**Coefficients of Variation (%) by Province, Territory and at the National Level for the Estimation of Average Household Expenditures for Several Summary Level Expenditure Categories and for the Estimation of Average Income**

| Summary level expenditure categories | Can. | N.L. | P.E.I. | N.S. | N.B. | Que. | Ont. | Man. | Sask. | Alta. | B.C. | Y.T. | N.W.T. | Nvt. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total expenditure | 0.7 | 1.6 | 1.9 | 1.2 | 1.2 | 1.2 | 1.4 | 1.2 | 1.4 | 1.4 | 1.2 | 5.7 | 4.6 | 8.1 |
| Total current consumption | 0.6 | 1.4 | 1.7 | 1.1 | 1.3 | 1.1 | 1.3 | 1.4 | 1.4 | 1.5 | 1.2 | 5.0 | 3.7 | 6.9 |
| Food | 0.5 | 1.1 | 1.6 | 1.1 | 1.1 | 0.9 | 1.1 | 1.2 | 1.3 | 1.2 | 1.2 | 3.7 | 2.8 | 4.7 |
| Shelter | 0.8 | 1.8 | 3.0 | 1.6 | 2.4 | 1.5 | 1.4 | 2.0 | 1.9 | 2.1 | 1.5 | 5.8 | 6.8 | 12.6 |
| Household operation | 0.9 | 2.3 | 4.1 | 1.8 | 2.2 | 1.8 | 1.8 | 2.2 | 2.2 | 2.0 | 1.8 | 6.3 | 6.9 | 7.8 |
| Furnishings | 2.4 | 3.5 | 5.5 | 3.4 | 4.1 | 3.7 | 4.7 | 3.9 | 3.6 | 5.8 | 4.4 | 9.1 | 5.2 | 10.5 |
| Clothing | 1.2 | 2.7 | 3.0 | 2.4 | 1.9 | 2.2 | 2.4 | 2.5 | 2.7 | 2.5 | 2.5 | 7.2 | 5.3 | 7.5 |
| Transportation | 1.4 | 3.2 | 4.7 | 3.2 | 3.3 | 3.1 | 2.8 | 3.8 | 3.6 | 3.1 | 3.1 | 8.0 | 4.6 | 22.8 |
| Health care | 1.3 | 3.1 | 6.4 | 2.6 | 3.0 | 2.1 | 3.5 | 2.7 | 2.9 | 2.2 | 2.7 | 10.4 | 10.6 | 31.3 |
| Personal care | 0.9 | 2.0 | 2.5 | 2.0 | 1.8 | 2.0 | 1.8 | 1.9 | 2.0 | 2.1 | 2.0 | 5.1 | 4.0 | 8.0 |
| Recreation | 1.8 | 3.3 | 3.9 | 3.3 | 3.4 | 3.1 | 3.6 | 4.1 | 3.3 | 4.4 | 4.3 | 7.7 | 6.8 | 9.6 |
| Reading & printed materials | 1.6 | 3.8 | 4.8 | 3.6 | 3.1 | 3.1 | 3.3 | 3.4 | 3.2 | 3.3 | 3.1 | 12.6 | 9.1 | 9.3 |
| Education | 3.1 | 8.3 | 10.9 | 10.8 | 8.0 | 5.5 | 5.6 | 8.5 | 7.9 | 5.7 | 6.4 | 19.4 | 24.9 | 36.3 |
| Tobacco, alcoholic beverages | 1.5 | 4.3 | 5.6 | 3.9 | 5.1 | 3.0 | 3.0 | 4.3 | 4.5 | 3.7 | 4.1 | 6.5 | 5.8 | 8.3 |
| Games of chance (net) | 3.4 | 10.4 | 12.9 | 8.4 | 7.6 | 6.8 | 6.5 | 10.7 | 13.6 | 9.8 | 8.6 | 22.4 | 16.6 | 17.6 |
| Miscellaneous expenditures | 3.0 | 6.9 | 8.2 | 6.4 | 7.5 | 6.2 | 6.2 | 6.1 | 5.2 | 5.1 | 6.2 | 18.9 | 10.4 | 34.0 |
| Personal income tax | 1.7 | 4.3 | 4.4 | 2.6 | 2.1 | 2.2 | 3.5 | 2.5 | 2.7 | 3.6 | 2.8 | 9.2 | 8.3 | 13.7 |
| Personal insurance | 1.5 | 2.4 | 3.1 | 2.5 | 1.7 | 2.7 | 3.1 | 1.9 | 2.4 | 2.6 | 2.5 | 12.5 | 6.1 | 8.5 |
| Gifts and contributions | 3.6 | 6.9 | 9.7 | 6.9 | 10.0 | 8.7 | 6.1 | 5.9 | 6.9 | 7.9 | 10.0 | 13.3 | 17.4 | 27.4 |
| Income | 0.7 | 1.4 | 1.7 | 1.0 | 0.8 | 1.3 | 1.4 | 0.9 | 1.1 | 1.8 | 1.0 | 8.5 | 6.8 | 11.4 |

The coefficients of variation (CVs) of the average estimates of total expenditure per household vary between 1.2% and 1.9% for the provinces and it is 0.7% at the national level. The CVs are higher in the territories, namely 5.8% in the Yukon, 5.7% in the Northwest Territories and 8.1% in Nunavut.

For summary level expenditure categories, the CVs at the national level are less than or equal to 2.4%, except for the following categories: *education*, *games of chance*, *miscellaneous expenditures* and *gifts of money and contributions*. These expenditure categories represent respectively 1.6%, 0.5%, 1.5% and 2.2 % of total expenditure (data not presented). Also, with the exception of these categories, the CVs are generally less than or equal to 6% at the provincial level.

Table 1.2 gives an overview of the CVs for some dwelling characteristics and household equipment estimates at the provincial and territorial level as well as at the national level.

**Table 1.2**
**Coefficients of Variation (%) at the National, Provincial and Territorial Level for Some Characteristics of Dwelling and Household Equipment**

| CATEGORIES | Can. | N.L. | P.E.I. | N.S. | N.B. | Que. | Ont. | Man. | Sask. | Alta. | B.C. | Y.T. | N.W.T. | Nvt. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Owner | 0.9 | 1.6 | 3.1 | 1.7 | 1.7 | 1.9 | 1.8 | 1.9 | 1.8 | 1.8 | 1.6 | 8.6 | 13.9 | 16.0 |
| Renter | 1.6 | 5.1 | 7.2 | 4.5 | 4.5 | 2.7 | 3.3 | 4.7 | 4.6 | 4.5 | 2.9 | 15.0 | 13.5 | 4.1 |
| Washing machine | 0.6 | 0.9 | 2.0 | 1.3 | 1.1 | 1.0 | 1.5 | 1.7 | 1.2 | 1.3 | 1.4 | 3.0 | 4.3 | 5.4 |
| Clothes dryer | 0.7 | 1.1 | 2.2 | 1.4 | 1.3 | 1.1 | 1.5 | 1.7 | 1.1 | 1.4 | 1.3 | 3.5 | 4.2 | 5.7 |
| Dishwasher | 1.1 | 4.3 | 5.1 | 3.1 | 3.2 | 2.1 | 2.5 | 2.8 | 2.9 | 2.2 | 2.0 | 11.9 | 13.1 | 33.6 |
| Freezer | 1.1 | 1.4 | 2.5 | 1.9 | 2.0 | 2.5 | 2.3 | 1.7 | 1.7 | 1.8 | 2.1 | 5.8 | 5.1 | 9.4 |
| Microwave oven | 0.4 | 0.9 | 1.2 | 0.7 | 0.7 | 0.9 | 0.7 | 0.7 | 0.9 | 0.6 | 0.9 | 2.9 | 2.9 | 6.7 |
| Cellular phone | 1.2 | 3.9 | 5.5 | 3.4 | 4.0 | 2.9 | 2.1 | 2.9 | 2.9 | 2.1 | 2.5 | 17.2 | 15.1 | 24.4 |
| CD player | 0.7 | 1.7 | 3.3 | 1.6 | 1.6 | 1.6 | 1.4 | 1.6 | 1.5 | 1.1 | 1.3 | 4.3 | 3.1 | 3.5 |
| Cable TV | 1.0 | 2.1 | 5.3 | 2.3 | 2.4 | 2.4 | 2.0 | 2.3 | 2.8 | 1.8 | 1.6 | 10.3 | 11.0 | 3.5 |
| Home computer | 0.9 | 2.6 | 4.3 | 2.3 | 2.7 | 2.2 | 1.7 | 2.3 | 2.6 | 1.7 | 1.7 | 8.3 | 5.5 | 14.0 |
| Regular tel. connection to a computer (Modem) | 1.9 | 5.1 | 7.0 | 4.6 | 4.0 | 3.8 | 3.4 | 5.4 | 5.6 | 4.9 | 4.4 | 12.4 | 11.3 | 19.2 |
| Use of internet (home) | 1.1 | 3.2 | 5.4 | 2.8 | 3.4 | 2.5 | 2.0 | 2.6 | 3.3 | 2.2 | 2.2 | 8.5 | 5.4 | 17.7 |
| Owned vehicles (one) | 1.5 | 3.1 | 4.7 | 3.2 | 3.4 | 2.9 | 3.2 | 3.2 | 3.1 | 3.2 | 2.6 | 8.3 | 6.3 | 14.7 |
| Owned vehicles (2 or more) | 1.4 | 4.2 | 4.6 | 3.4 | 3.7 | 3.3 | 3.0 | 3.2 | 2.6 | 2.4 | 2.5 | 9.1 | 16.4 | 37.2 |

The coefficients of variation for the estimates of dwelling characteristics and household equipment are generally below 5% at the provincial level, with a few exceptions in the category *Regular telephone connection to a computer (Modem)*. The CVs are higher in the territories where there is a smaller proportion of equipment.

The coefficients of variation for the dwelling characteristics and the household equipment at the national level are below or equal to 1.9%.

## 1.3 Model for Deriving an Approximation of the CV

Estimates for different domains of interest (for example, by income quintile) for the summary level expenditure categories are available in the publication *Spending Patterns in Canada* (see reference [4]). Estimates for different domains of interest for the detailed expenditure are available upon request from the Income Statistics Division. (For more details on tables available upon request from the Income Statistics Division, see references [3] or [4].) For operational reasons, it is not possible to produce CVs for all the characteristics collected by the survey at all the different levels of aggregation that may interest users.

### *1.3.1 Model for Deriving an Approximation of the CV for Domain Estimates*

It is however possible to calculate an approximation of the CV by using a relationship between the number of households in the sample that reported expenditures for a given category and the CV at an aggregated level. This relationship, based on the CV's tendency to increase in proportion to a decrease in the square root of the number of households reporting an expenditure, is illustrated below.

***Formula for Approximating the CV for a Domain (Subgroup of the Population)***

If CV(Y) represents the CV for the estimate of the average per household of a certain characteristic for the entire population, then an approximation of the CV of the estimate of that characteristic can be calculated for a domain (which may be considered as a subgroup of the population, such as a household type, an income quintile, an urbanization level, etc.) according to the following equation:

$$CV(Y_d) = CV(Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

where

*n:* *number of households in the sample*
*P:* *estimate of the proportion of households reporting a value > 0 for this characteristic in the population*
$n_d$: *number of households in the sample in domain d*
$P_d$: *estimate of the proportion of households reporting a value > 0 for this characteristic in domain d*

Generally, approximations for the different domains are calculated using the CV, size *n* and proportion *P* at the national level. If an approximation of a CV is desired for a domain that is entirely contained within a single province (for example, a metropolitan area), then it is preferable to use the values at the provincial level, since provincial CVs are published for the 2001 SHS (reference [3]). It should be noted that a CV obtained using this approach is only an approximation of the real value.

### *1.3.2 Method of Computation of an Approximate CV from the Microdata File*

The microdata file users can obtain an approximation of the CV of the estimates using another method which will generally provide better results than the method described in the previous section for the CVs of detailed expenditure estimates. This approach is described in detail in the documentation provided with the 2001 microdata file. This method of approximation can only be used with the microdata file since it is necessary to have data and weights for each household.

The document on data quality for the 1997 SHS contains the results from the performance evaluation of these two CV approximation methods.

### 1.4 Suppression of Unreliable Data in Estimation Tables

Since the coefficient of variation is an indicator of the reliability of data, we would like to use it to determine whether or not the estimates should be published. Estimates for which the CV is more than 33% are not considered sufficiently reliable to be published. However, CV estimates are not calculated for many of the published estimates. The suppression rule for expenditure estimates is therefore based on the number of households reporting a value greater than zero.[3]

It can be shown that CVs are usually below 33% when the number of households reporting an expenditure is greater than 30. Since this is an approximate rule, some estimates may be published even though the CV is greater than 33%, and some estimates will not be published even though the CV is less than 33%. The document on data quality for the 1997 SHS gives the results from the evaluation of the risk of error in the use of the suppression rule.

## 2. Nonresponse

Errors due to nonresponse result from the fact that some potential respondents do not provide the necessary information or the provided information proves to be unusable. When the respondent has failed to respond to only some questions, this is referred to as partial nonresponse. In such a case, the missing data are imputed. Errors associated with imputation are described in Section 5, which deals with processing errors. In the present section, nonresponse includes collection nonresponse, which is mainly due to the inability to contact the household or to the refusal of the members of the household to participate partially or completely in the survey, as well as data collected from households that prove to be unusable.

The main impact of nonresponse on data quality is that it can introduce a bias in the estimates if the characteristics of respondents and nonrespondents differ and the difference has an impact on the characteristics studied. Nonresponse rates may easily be calculated, but they have only an indicative value with regard to data quality, since they do not allow estimation of the bias associated with the estimates. The scope of nonresponse may be considered as an indicator of the risks of bias in the estimates.

### 2.1 Nonresponse Rates and Vacancy Rates

Since the units selected in the SHS are dwellings, interviewers must first identify ineligible dwellings, that is, dwellings occupied by persons who are not part of the target population, as well as dwellings that no longer exist (demolished, mobile home moved or dwelling converted to business) and vacant dwellings (unoccupied, seasonal or under construction).

Among eligible dwellings, the proportion of households that did not respond to the survey is evaluated next. This is called the collection nonresponse rate. Included are households that refused to participate in the survey and households where no contact

[3] In practice, we use the estimate of the proportion of households reporting an expenditure, which is multiplied by the sample size.

could be made with the respondents, either because they were absent or because of special circumstances (language problem, illness, death).

Again among eligible dwellings, the rate of unusable data is determined. Unusable data refers to the number of households whose questionnaires were at least partially completed but which were rejected during data processing. There are two main causes for rejection. First, when many questions on income or expenditures have been left unanswered, the questionnaire is classified as incomplete and is not used. The other source of rejection consists of questionnaires in which the difference between receipts (income and other sources of money received by the household) and disbursements (expenditures and net change in assets and liabilities) is greater than 20%. These questionnaires are also excluded from the estimation and are considered as nonresponse.

Table 2.1 shows the final nonresponse rate as well as the collection nonresponse rate, broken down by refusals and no-contacts, and the rate of unusable data broken down into incomplete and unbalanced questionnaires. The vacancy rate is also included. These rates are provided at the national level as well as at the provincial level.

**Table 2.1**
**Nonresponse Rates (%) and Vacancy Rates (%) by Province, Territory and at the National Level**

| Province or territory | Vacancy rate | Collection nonresponse rate | | | Unusable data rate | | | Final nonresponse rate (at estimation stage) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | No contact | Refusal | TOTAL | Incom-plete | Un-balanced | |
| **Canada** | 11.5 | 20.4 | 5.2 | 15.2 | 3.4 | 1.9 | 1.5 | 23.8 |
| Newfoundland and Labrador | 18.4 | 16.1 | 5.4 | 10.8 | 2.1 | 1.1 | 1.0 | 18.3 |
| Prince Edward Island | 19.3 | 15.3 | 4.6 | 10.7 | 0.8 | 0.4 | 0.4 | 16.0 |
| Nova Scotia | 13.5 | 21.6 | 6.7 | 14.9 | 4.3 | 3.0 | 1.3 | 25.9 |
| New Brunswick | 15.9 | 18.4 | 4.5 | 14.0 | 2.8 | 1.4 | 1.4 | 21.2 |
| Quebec | 6.9 | 21.1 | 6.5 | 14.6 | 1.5 | 0.9 | 0.6 | 22.7 |
| Ontario | 7.7 | 28.1 | 7.3 | 20.8 | 7.7 | 4.8 | 2.9 | 35.8 |
| Manitoba | 13.8 | 15.4 | 3.3 | 12.1 | 1.8 | 0.6 | 1.2 | 17.2 |
| Saskatchewan | 10.8 | 17.5 | 4.1 | 13.4 | 2.9 | 1.2 | 1.7 | 20.4 |
| Alberta | 5.3 | 19.3 | 3.2 | 16.1 | 3.5 | 1.5 | 2.0 | 22.8 |
| British Columbia | 7.7 | 22.3 | 3.8 | 18.5 | 3.1 | 1.8 | 1.3 | 25.4 |
| Yukon | 31.4 | 24.1 | 3.8 | 20.3 | 3.2 | 2.3 | 0.9 | 27.2 |
| Northwest Territories | 13.3 | 11.7 | 6.1 | 5.6 | 1.5 | 0.7 | 0.7 | 13.1 |
| Nunavut | 12.6 | 15.0 | 7.7 | 7.3 | 0.4 | 0.4 | 0.0 | 15.4 |

The final nonresponse rate at the national level is 23.8%. It is due to refusals (15.2%), to households that could not be contacted (5.2%), and finally to households for which the data were unusable (3.4%). For most provinces, refusals are the main cause of nonresponse, followed by the households that could not be contacted, and by the households for which the data were unusable. The only exceptions are Ontario and Alberta where the proportion of unusable data is slightly higher than the proportion of households that we were unable to contact.

The final nonresponse rate varies from one province to another. The Northwest Territories has the lowest nonresponse rate at 13.1%. The lowest refusal rate (5.6%) is also observed in the Northwest Territories. Nonresponse rates in Newfoundland and Labrador, Prince Edward Island, Manitoba and Nunavut are smaller than 20% while rates over 20% are observed for the other provinces and territories. Ontario has the highest nonresponse rate at almost 36%. The highest refusal rate (20.8%) is also observed in Ontario.

The vacancy rates are shown in Table 2.1, but it should be kept in mind that vacant dwellings do not contribute to the bias of the sample if they are correctly identified. By analysing vacancy rates, we can detect dwelling identification problems associated with the collection process. The national vacancy rate for the 2001 SHS is 11.5%, which is slightly lower than for the Labour Force Survey (LFS) for the same period. Note that the higher vacancy rate (31.4%) observed in the Yukon is due in part to a different sample allocation for 2001 for this territory and to a lesser extent to modifications made by the LFS to the coverage in the Yukon.

## 2.2 Nonresponse According to Urbanization Level

Nonresponse varies according to urbanization level. The various rates at the national level are shown by urbanization level in Table 2.2.[4]

**Table 2.2**
**Nonresponse Rates (%) and Vacancy Rates (%) by Urbanization Level**

| Urbanization category | Vacancy rate | Collection nonresponse rate | | | Unusable data rate | | | Total nonresponse rate (at estimation stage) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | No contact | Refusal | TOTAL | Incomplete | Unbalanced | |
| **URBAN** | | | | | | | | |
| 1,000,000 or more | 4.1 | 26.8 | 7.8 | 19.0 | 3.7 | 2.1 | 1.6 | 30.6 |
| 500,000 to 999,999 | 3.9 | 18.7 | 3.3 | 15.3 | 2.7 | 1.0 | 1.7 | 21.4 |
| 250,000 to 499,999 | 6.3 | 26.0 | 6.9 | 19.2 | 5.9 | 3.8 | 2.2 | 32.0 |
| 100,000 to 249,999 | 7.9 | 22.0 | 5.0 | 17.0 | 3.3 | 2.1 | 1.1 | 25.3 |
| 30,000 to 99,999 | 5.3 | 19.1 | 3.7 | 15.4 | 3.3 | 2.1 | 1.1 | 22.3 |
| Less than 30,000 | 12.5 | 17.1 | 4.6 | 12.5 | 3.0 | 1.6 | 1.4 | 20.1 |
| **RURAL** | 23.4 | 16.6 | 4.6 | 12.0 | 3.2 | 1.8 | 1.4 | 19.8 |
| **TOTAL** | 11.5 | 20.4 | 5.2 | 15.2 | 3.4 | 1.9 | 1.5 | 23.8 |

The final nonresponse rate generally increases with urbanization level. According to Table 2.2, the "500,000 to 999,999" group and, to a lesser extent, the "1,000,000 or more" group, go against this rule.

The collection nonresponse rate also increases with urbanization level. There is a big difference of nearly 10% between the urbanization categories "less than 30,000" and "1,000,000 or more". Refusals account for more than 60% of the total nonresponse at each level of urbanization. Although the urbanization category "500,000 to 999,999" has

[4] Tables on nonresponse rates by urbanization level and province are available on request from the Household Survey Methods Division.

the lowest total nonresponse rate among urban regions of 30,000 persons or more, this category has the highest proportion of nonresponse (71.5%) due to refusals.

From an examination of the vacancy rate by urbanization level, it appears that the vacancy rate at the rural level (23.4%) is nearly more than two times higher than for low-population urban areas (12.5%). The latter also shows a rate that is higher than for the high-population urban areas. This phenomenon is also observed in the LFS and may be explained by a greater number of seasonal dwellings in rural areas. Since the SHS sample is more concentrated in high-population urban areas than the LFS, the national vacancy rate for the SHS can be expected to be slightly lower than that for the LFS.

## 2.3 Nonresponse According to Income Strata

Since income information is not available for nonrespondents, it is not possible to compare nonresponse rates according to income. However, the LFS sample design, used for the SHS, was designed in such a way that in nine large cities there are strata consisting of geographic areas where the average household income exceeds $100,000, and in seven large cities there are strata consisting of apartments inhabited by households with an average income of less than $20,000. Even though the number of such strata is small and accounts for only a small number of dwellings in the SHS sample (approximately 580 for high income and 160 for low income strata, or less than 3% of the sample), the comparison of nonresponse rates for these two groups in relation to the other strata is revealing. Table 2.3 shows these results.

**Table 2.3**
**Comparison of Nonresponse and Vacancy Rates (%) in High-income and Low-income Strata in Relation to Other Strata**

| Stratum type based on income | Vacancy rate | Collection nonresponse rate | | | Unusable data rate | | | Final nonresponse rate (at estimation stage) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | No contact | Refusal | TOTAL | Incom-plete | Un-balanced | |
| High-income | 2.9 | 31.5 | 8.8 | 22.7 | 5.9 | 1.8 | 4.1 | 37.5 |
| Regular | 11.7 | 20.1 | 5.0 | 15.1 | 3.3 | 1.9 | 1.4 | 23.5 |
| Low-income | 5.1 | 17.9 | 6.2 | 11.7 | 0.7 | 0.7 | 0.0 | 18.6 |
| TOTAL | 11.5 | 20.4 | 5.2 | 15.2 | 3.4 | 1.9 | 1.5 | 23.8 |

The final nonresponse rate (37.5%) in high-income strata is much higher than in regular strata. The refusal rate for high-income strata is over 20%, which is nearly twice the rate of low-income strata.

Households in low-income strata have a collection nonresponse rate similar to those for the regular strata. However, regular strata have a final nonresponse rate of 26% higher than that for low-income strata. This is mainly explained by the fact that there are almost no unusable data in the low-income strata.

As for the 1997 to 2000 SHS, the vacancy rate is higher for regular strata than for each of the other two strata. In high-income and low-income strata, the vacancy rates are almost the same.

## 2.4 Adjustment for Nonresponse

To compensate for nonresponse, the weights in the SHS are inflated by the inverse of the weighted response rate within certain groups defined on the basis of the different urbanization levels in each province. The weighted rates differ from the rates presented in this section since the former takes into account the sampling weight of each household. An algebraic description of the nonresponse adjustment is provided in Appendix A.

The weights adjustment for nonresponse takes into account the differences in nonresponse by urbanization level as described in Section 2.2. It will reduce the bias to the extent that the characteristics of respondents and nonrespondents are similar for a given urbanization level.

# 3. Coverage errors

In the design of the survey, the target population was defined. It is useful to go over this definition, since a good understanding of the target population is necessary in order to properly interpret the survey data. One should note that SHS uses the LFS sampling frame.

*Target population*

> The target population consists of individuals living in private households. It therefore excludes residents of institutions such as prisons, chronic care hospitals or senior citizens' homes, as well as members of religious orders and other groups living communally, members of the Armed Forces living in military compounds, and individuals residing permanently in hotels or rooming houses. Also excluded are foreign countries' official representatives residing in Canada and their families as well as individuals residing on Indian reserves or public lands. With these exclusions, the survey covers nearly 98% of the population in the 10 provinces. In the Yukon, the coverage of the population went from 81% in 1999 to 88% in 2001. This increase is due to modifications made by LFS to the coverage of this territory. Thus, some communities and unorganized areas that were previously excluded are now included. The coverage of the Northwest Territories represents 92% of the population and the coverage of the Nunavut represents 89% of the population.[5]

Coverage errors result from inadequate representation of the target population based on the units of the sampling frame. Some units of the target population may be omitted from the sampling frame, in which case there is undercoverage. Other units that are not in the target population may be included by error, or some units may be included more than once. These units are responsible for overcoverage.

[5] In terms of households, the coverage of the Yukon, the Northwest Territories and Nunavut represents respectively 88%, 93% and 90% of households.

## 3.1 Undercoverage and Overcoverage: Slippage Rates

In the SHS, the sample is selected using a list of dwellings in each selected cluster. Factors contributing to undercoverage are: the omission of dwellings in the creation of the list, new dwellings that are added between the creation of the list and the interviewer's visit (mainly in developing areas), and the erroneous classification of vacant dwellings. The inclusion of dwellings that are not within the boundaries of the cluster is a source of overcoverage. Similarly, errors can take place during data collection, due to improper identification of persons as members of the selected household. These errors also contribute to undercoverage or overcoverage.

A good representation of the target population is essential to the production of realistic expenditure estimates. The number of people per household is also an important characteristic in the estimation of average household expenditures. Therefore, it is necessary that the sample not only adequately represent the individuals in the target population, but also the distribution of households according to their size.

In 1999, a weighting strategy that uses new controls was introduced. This method results in a better correction of the representation of the target population by using a more detailed age-sex grouping than was used previously and for which the coverage varies from one group to the other.

There is generally a net undercoverage of the number of persons in the SHS. This undercoverage is corrected by an adjustment of weights using auxiliary data based on post-censal demographic estimates. The slippage rate (see appendix A) is a measure of the percentage of difference between the auxiliary data and the survey estimates calculated using weights not adjusted with these data.[6] Tables 3.1 and 3.2 respectively show slippage rates by age-sex group at the national level and at the provincial level. Table 3.3 presents these rates for the household size categories used for the weight adjustment. A positive rate indicates overcoverage of the number of persons in the survey.

[6] The subweight which is the survey weight adjusted for nonresponse is used (see Appendix A).

**Table 3.1**
**National Slippage Rates by Age-Sex Group**

| National Slippage Rates by Age-Sex Group | | | | |
|---|---|---|---|---|
| | Age | Sex | | Total |
| | | Male | Female | |
| Canada[1] | 0-6 years | 0.5 | 5.2 | 2.8 |
| | 7-17 years | -3.5 | -1.6 | -2.6 |
| | 18-24 years | -6.5 | -3.2 | -4.9 |
| | 25-34 years | -13.4 | -10.2 | -11.8 |
| | 35-54 years | -13.6 | -7.9 | -10.7 |
| | 55-59 years | -7.3 | -15.9 | -11.7 |
| | 60-64 years | -12.4 | -4.7 | -8.4 |
| | 65-69 years | -6.4 | -0.8 | -3.5 |
| | 70 years and + | -9.4 | -2.5 | -5.4 |
| | **Total** | **-9.2** | **-5.4** | **-7.3** |

[1] Excluding the territories.

For the 2001 SHS, the national undercoverage rate was 7.3%. If we analyze Table 3.1 with respect to age group, we can see that national slippage rates for children (0 to 6 and 7 to 17) are quite different from those of other age groups. Overcoverage or a slight undercoverage occurs with respect to children, while there is always undercoverage among adults. The undercoverage rate for all children combined is 0.7%, while it is 9.2% for adults (data not presented).

The highest national rates occurred among 55 to 59 years old women, and among 25 to 34 years old and 35 to 54 years old men. Except for the category 55 to 59 years old, the undercoverage rate for women is always smaller than the undercoverage rate for men.

As mentioned previously, the SHS uses the LFS sampling frame. Over the same period, the national LFS undercoverage rate was 9.8% (reference [5]). This is slightly higher than the 8.8% SHS rate for those 15 years or older (data not presented).

**Table 3.2**
**Slippage Rates for Provinces and Territories by Age-Sex Group**

| Slippage Rates by Age-Sex Group | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sex | Age | Newfoundland and Labrador | Prince Edward Island | Nova Scotia | New Brunswick | Quebec | Ontario | Manitoba | Saskatchewan | Alberta | British Columbia | Yukon | Northwest Territories | Nunavut |
| Male | 0-6 | -0.1 | -21.1 | -2.0 | -8.1 | 12.0 | 0.2 | -14.8 | 6.4 | -8.3 | -4.7 | 7.4 | -18.3 | -19.4 |
| | 7-17 | -17.4 | 5.1 | -3.3 | -7.4 | -10.9 | 7.3 | -6.8 | 1.6 | -8.1 | -18.6 | | | |
| | 18-24 | -27.1 | -31.6 | -18.0 | -23.0 | -1.5 | 1.9 | -13.9 | -31.6 | -21.9 | -9.0 | | | |
| | 25-34 | -43.3 | -13.8 | -23.8 | -25.9 | -7.8 | -9.6 | -17.2 | -12.7 | -13.7 | -25.6 | | | |
| | 35-54 | -14.0 | -15.5 | -11.9 | -15.5 | -12.5 | -11.9 | -11.3 | -17.5 | -13.7 | -19.9 | | | |
| | 55-59 | -18.6 | -11.3 | -15.9 | -6.0 | -1.3 | -8.8 | -11.8 | -24.3 | -1.5 | -10.2 | -18.1 | -27.4 | -20.3 |
| | 60-64 | -0.3 | -6.5 | -12.1 | -38.1 | -7.7 | -15.7 | -16.4 | -28.5 | -11.5 | -4.8 | | | |
| | 65-69 | 15.5 | -15.5 | 15.2 | 4.5 | -5.4 | -11.7 | -10.8 | -13.1 | 2.5 | -5.1 | | | |
| | 70 + | 9.2 | -3.9 | -11.6 | -4.6 | -12.7 | -12.7 | -0.3 | 2.0 | -1.9 | -7.3 | | | |
| | **Total** | **-16.1** | **-12.6** | **-11.2** | **-14.8** | **-7.5** | **-6.3** | **-11.3** | **-11.9** | **-11.4** | **-15.5** | **-11.5** | **-24.6** | **-19.9** |
| Female | 0-6 | -21.7 | 39.3 | -1.6 | 0.8 | 14.5 | 3.8 | 7.9 | 8.6 | 4.9 | -4.3 | 10.3 | -15.8 | -22.8 |
| | 7-17 | -9.9 | -1.9 | 0.4 | -4.5 | -4.2 | 8.3 | -10.4 | -2.6 | -10.8 | -16.2 | | | |
| | 18-24 | -30.4 | -34.8 | -18.5 | -16.3 | 14.7 | -5.2 | -14.9 | -24.6 | 5.6 | -17.7 | | | |
| | 25-34 | -27.3 | -8.1 | -15.9 | -11.8 | -2.1 | -11.7 | -2.0 | 0.4 | -11.9 | -18.4 | | | |
| | 35-54 | -11.6 | -12.0 | -6.3 | -5.9 | -9.9 | -3.0 | -7.5 | -7.4 | -8.7 | -17.8 | | | |
| | 55-59 | -8.0 | 4.5 | -10.1 | -10.5 | -2.1 | -23.3 | -27.6 | -28.2 | -17.6 | -20.7 | -10.9 | -19.0 | -13.3 |
| | 60-64 | 12.4 | -9.9 | 15.9 | 2.2 | -7.0 | -7.3 | -12.8 | -5.6 | -1.0 | -1.0 | | | |
| | 65-69 | -14.3 | -23.1 | -9.7 | 6.9 | 2.8 | -4.4 | 14.7 | -9.9 | 13.8 | -5.0 | | | |
| | 70 + | 1.9 | 6.2 | -10.1 | -12.1 | -8.2 | 1.6 | -2.4 | -1.0 | 3.5 | -4.3 | | | |
| | **Total** | **-13.7** | **-6.0** | **-7.2** | **-7.0** | **-2.8** | **-3.2** | **-6.5** | **-6.2** | **-5.3** | **-14.3** | **-5.7** | **-17.9** | **-17.4** |
| **Total** | | **-14.9** | **-9.3** | **-9.1** | **-10.9** | **-5.2** | **-4.7** | **-8.9** | **-9.1** | **-8.4** | **-14.9** | **-8.6** | **-21.3** | **-18.7** |

We observe a net undercoverage for all provinces and territories, with the rates varying from 4.7% to 21.3%. The magnitude of the undercoverage is greater for the Northwest Territories and Nunavut. However, a low overall rate of undercoverage is not a guarantee of a better coverage. For example, Prince Edward Island's overall slippage rate (-9.3%) concealed the worst case of overcoverage for a provincial age-sex group (39.3% among 0 to 6 years old females) and the third and fourth worst cases of undercoverage (34.8% and 31.6% among 18 to 24 years old women and men respectively). On the other hand, despite an overall slippage rate of -14.9%, British Columbia has only the thirteenth worst case of undercoverage (25.6% among 25 to 34 years old men). We can also see that the pattern of slippage rates differs substantially for age-sex groups from one province to the next.

**Table 3.3**
**Slippage Rates for Provinces and Territories by Household Size**

| Province or Territory | Slippage Rate | | | |
|---|---|---|---|---|
| | Households | One-person households | Two-person households | Three-person and more households |
| **Canada** | **-6.0** | **-4.0** | **-6.1** | **-7.3** |
| Newfoundland and Labrador | -10.1 | 0.8 | -0.3 | -20.3 |
| Prince Edward Island | -6.3 | -4.6 | 6.2 | -16.2 |
| Nova Scotia | -7.4 | -2.8 | -5.8 | -11.4 |
| New Brunswick | -6.7 | 0.2 | 1.0 | -16.4 |
| Quebec | -5.3 | -10.9 | 0.0 | -5.4 |
| Ontario | -4.9 | 0.0 | -12.9 | -2.1 |
| Manitoba | -4.4 | 4.8 | -6.0 | -9.6 |
| Saskatchewan | -7.0 | -4.5 | -2.0 | -13.1 |
| Alberta | -4.1 | -0.7 | 2.6 | -10.9 |
| British Columbia | -11.0 | -4.1 | -9.5 | -16.8 |
| Yukon | -12.2 | -32.1 | -3.9 | -4.7 |
| Northwest Territories | -21.6 | -17.9 | -5.0 | -29.6 |
| Nunavut | -17.7[7] | .. | .. | .. |

Nationally, the number of households was underestimated by 6%. This slippage rate is comparable to the -7.3% slippage rate for individuals.

Among the provinces and territories, there is a substantial variation in the slippage rate for one-person households. This rate ranges from -32.1% for the Yukon to 4.8% in Manitoba. The same phenomenon occurs with respect to two-person and three-person and more households but with less variation. For two-person households, the slippage rate ranges from -12.9% in Ontario to 6.2% in Prince Edward Island. The slippage rate for households of three or more persons ranges from -29.6% in the Northwest Territories to -2.1% in Ontario. For households of three or more persons, there is always undercoverage. Excepting Quebec, Ontario and the Yukon, the greatest undercoverage is observed for households of three or more persons.

## 3.2 Adjustment for Coverage Error at the Population Level and Household Level

To correct the coverage problem illustrated in Tables 3.1 and 3.2 and to reduce the resulting bias, the survey data are adjusted during weighting using demographic estimates for the age groups defined in these tables, for each province. For more information on the adjustment's methodology, see reference [1]. This adjustment greatly reduces the bias caused by coverage errors but does not completely eliminate bias if the characteristics of the individuals omitted from the survey differ from those of individuals included for a given age group in a province.

Furthermore, the effectiveness of the coverage adjustment based on demographic estimates depends mostly on the quality of those estimates and their accuracy in

[7] Only the total number of households was used for Nunavut.

representing the survey's target population. The demographic estimates are not error-free. They are post-censal estimates based on the population counts from the 1996 Census adjusted for net undercoverage, and they take into account recent statistics on migration, births, deaths, etc. These demographic estimates are adjusted to account for certain exclusions specific to household surveys, such as persons living in institutions. Conceptually, they differ slightly from the SHS target population in that they include persons living in non-institutional collective dwellings, such as members of groups living communally and individuals permanently residing in hotels or rooming houses. However, this difference is considered negligible, since such individuals represent less than 0.4% of the Canadian population.

To remedy the issue of the sample's representativity with respect to the number of households based on their respective sizes as illustrated in Table 3.3, we use supplementary data to adjust the data appearing in the survey. By adjusting the weights of the SHS to reflect post-censal estimates of the number of households by size, we hope to compensate for the bias produced by inadequate representation of households. However, we will not necessarily succeed in eliminating such bias if features of uninterviewed (omitted or non-respondent) households differ from those of responding households for the same size or group. As in the case of demographic estimates of population, the success of such an adjustment depends on the quality of the supplemental data on the number of households. As such, note that some changes were made to the methodology used in the production of the supplemental data on the number of households for the Northwest Territories and Nunavut. The impact that these changes could have on the outcome of statistical tests comparing estimates from both 1999 SHS and 2001 SHS surveys has not been studied for these two territories.

In addition to demographic estimates of age-sex groups by province, three other groups of supplementary data are used during weighting to adjust survey data and thereby improve their representativity. The first set of data is used to control for the number of children and adults in certain major cities. The second is designed to control for the number of single-parent households and couples with children by province. Finally, counts for major categories of income from wages and salary are used when adjusting weights to ensure a certain degree of consistency between the income distributions from the SHS and from outside sources.

## 4. Response errors

Response errors represent a lack of accuracy in responses to questions. They can be attributed to different factors, including a questionnaire that requires improvements, misinterpretation of questions by interviewers or respondents, and errors in respondents' statements.

In the SHS, there can be various reasons for errors in respondents' statements. First, there are recall errors that occur when a respondent forgets expenditures made during the period covered by the survey (which corresponds to the calendar year), or when a respondent provides an erroneous value because of the time interval that has elapsed between the time of purchase and the date of the interview. Recall errors are probably the survey's largest source of response error, since the reference period is long (12 months) and a great variety of information is requested.

One of the main measures taken to minimize recall error in the SHS is to calculate the difference between receipts (income and other amounts received by the household) and disbursements (expenditures plus net change in assets and liabilities) for each household. When the difference exceeds 10% of receipts or disbursements, with the higher amount being retained, respondents are contacted again in order to obtain additional information and to try to identify errors or omissions. The respondent is also encouraged to consult various documents (invoices, bank statements, etc.) in order to provide more accurate data. To determine expenditures for small items purchased at regular intervals, interviewers generally suggest that respondents estimate the frequency of the purchases and the price generally paid in order to derive expenditures for a 12-month period.

A second source of error in respondents' reporting is telescopic error, which consists of including in the reference period events that occurred before or after it. In the SHS, the use of the calendar year is considered to provide a good marker for the start of the reference period. Furthermore, since the reference period is a long one, telescopic error has less impact.

Responses by proxy can also contribute to response error. The household member who made an expenditure is generally best able to report it accurately. This is definitely the case with, say, personal purchases. Expenditures reported by an intermediary are more likely to be tainted by response error, and this type of error tends to have a greater effect on certain types of expenditures.

Among other sources of response error, the extent of the respondent's co-operation should not be overlooked. For personal reasons, the respondent may decide not to mention particular expenditures or decide to twist the facts.

In the SHS, another factor is the response burden, owing to the length of the interview and the great variety of items to be reported, as well as the pace of the interview. This can lead to respondent fatigue and affect the quality of the responses obtained. The interview time varies greatly from one household to another, depending on household size, income and various other characteristics. For some households, the interview can take more than five hours.

It should be stressed that questions were added to the 2001 SHS. Among other things, for 2001 only, extra questions were included in the survey so that data from the SHS could be used in the weighting of the Consumer Price Index. This change may affect historical comparisons for a few variables. For example, questions were added under "Personal care preparations" to collect specific information about hair care products, makeup, fragrances, deodorants and oral hygiene products. As a result of these extra questions, respondents may have given more precise information and the increase in the estimate for "Personal care preparations" may have been at least partly caused by an improvement in respondent recall. The effect of additional questions on estimates is difficult to quantify.

While response errors are a major source of error in a historical interview, they are the aspects of data quality that are the hardest to measure. Generally, it is necessary to conduct quite costly special studies in an attempt to measure them. Efforts are made to combat response errors by using survey techniques designed to reduce them.

## 5. Processing errors

Processing errors can arise in all types of data handling. The main stages of data processing are coding, data entry, editing, imputation of partial nonresponse and weighting. In the SHS different procedures are applied at each stage in order to minimize processing errors and the survey estimates are compared with other data sources prior to release. Errors related to the adjustments made at the weighting stage have been described in sections 2 and 3. The other types of processing errors are covered in this section.

Coding is necessary for only a few questions. This is done by the interviewer and subsequently verified by a senior interviewer. Before 2001, data entry was done with the help of an automated verification system that grouped the questionnaires into batches and chose some questionnaires from each batch to be entered a second time. Any errors found were to be corrected. If the number of errors in a batch was greater than a certain threshold, then the entire batch was submitted for re-entry. Due to the introduction of a new data capture system (BLAISE), there was no questionnaire batch verification procedure this year. However, some edits were implemented in the new data capture system to ensure consistency of data captured. The results of a preliminary data capture study seem to show that data capture error rates of the new system are similar to the ones of the old system.

The first stage of automated verification is done after each questionnaire has been verified manually by both the interviewer and the senior interviewer. It is ensured that the respondent's answers follow some essential consistency rules. Unusual situations that may justify corrections are also identified. This stage of verification is done in Statistics Canada's regional offices in case it is necessary to recontact respondents if some supplementary information is required to resolve inconsistencies in the answers provided. Specially-trained members of the verification teams solve any problems identified. Thereafter, other verification checks are done at the head office and invalid responses are corrected.

The processing of SHS data also involves imputation for partial nonresponse. Partial nonresponse occurs when the respondent refuses to answer or does not know the answer to certain questions. The imputation approach differs depending on whether the data is categorical or continuous. Categorical data takes on only specific values (as in yes/no questions or type of dwelling questions), while continuous data can take any numerical value (as for income and expenditure data).

Categorical data, which are obtained mainly in the facilities and equipment section of the questionnaire, are imputed with the help of a "hot deck" imputation technique that randomly chooses a donor from a group of answering households with similar characteristics.

Income and expenditure data are imputed with the nearest neighbour technique. The imputation is done on one group of variables at a time, with the groups chosen by taking the relationships among the variables into account. A group generally corresponds to a section of the questionnaire. For every group, the missing values of a recipient (a household that has some missing data for at least one of these variables) are imputed from data of the most similar record among all donors (households that have no missing

values for these variables). For each recipient the closest donor is chosen as the one that minimizes a particular distance function. This function is based on matching variables that are chosen because they are correlated with the variables to be imputed. For example, the total income of a household is chosen as a matching variable for all sections pertaining to expenditures. It must also be ensured that, after receiving the donor values, the recipient household satisfies some consistency rules. In general, the imputation is done at the household level, but in some groups (e.g., income and clothing expenditures), the imputation is done at the person level since the original data is collected at that level.

Note that in 2001, the imputation of all expenditure and income data was done using a new system based on slightly different methodology from the one used by the previous system. The new system allows a better use of categorical variables as matching fields when selecting a donor. The new system was tested prior to its implementation and the results it gave were similar to those with the old system. Also, results with the new system are similar to those of previous years.

The bias caused by imputation of partial nonresponse is difficult to evaluate. It depends on the differences between respondents and nonrespondents as well as the ability of the imputation method to produce unbiased estimates. However, the imputation rates indicate the importance of partial nonresponse. They are presented in the following section.[8]

[8] For operational reasons, these data quality indicators are not available for categorical data such as Household Facilities and Equipment.

## 5.1 Proportion of Households or Persons Requiring Imputation at National, Provincial and Territorial Level

A preliminary indication of the magnitude of partial nonresponse is the proportion of households requiring imputation and the number of variables imputed by household. The questionnaire can be divided into two major groups of variables: those collected at the household level and those collected at the individual level (such as income and clothing expenditure). For the latter, the respondent may provide only the total income or total clothing expenditures if he/she is unable to provide the breakdowns by source of income or type of expenditure. The level of imputation for the components of income and clothing expenditure is then larger, but this does not affect the total income, total clothing expenditure or total expenditure.

The percentage of households requiring imputation for household expenditure (excluding clothing expenditures) is presented in the next sub-section. The following sub-section presents the percentage of persons requiring imputation for a clothing expenditure variable and the percentage of persons requiring imputation for an income variable. After data imputation by the system, some corrections might have been needed on both imputed and non-imputed variables aiming to ensure data consistency. The actual percentages of records requiring this kind of modification are about 1% at the household level and 2% at the person level. The results are provided at the national, provincial and territorial levels. This gives an indication of which provinces are more affected by imputation.

### *5.1.1 Household Expenditure Imputation by Province and Territory*

The percentage of usable households that required imputation for an expenditure variable (excluding clothing expenditures) is presented in Table 5.1. Usable households correspond to all households living in eligible dwellings, excluding households who could not be contacted, who refused to participate in the survey, or who provided unusable data (see section 2.1). The table is broken down by the number of imputed variables (out of 251) for a household.

**Table 5.1**
**Households Requiring Expenditure Imputation by Province and Territory**

| Province | Households (%) requiring imputation for EXPENDITURE VARIABLES (excluding clothing expenditures) | | | |
|---|---|---|---|---|
| | Number of variables imputed (out of 251) | | | |
| | 1 | 2 | 3 or more | TOTAL |
| **Canada** | **5.4** | **1.0** | **0.8** | **7.2** |
| Newfoundland and Labrador | 9.9 | 0.8 | 0.2 | 11.0 |
| Prince Edward Island | 2.9 | 0.5 | 0.8 | 4.1 |
| Nova Scotia | 7.0 | 1.2 | 1.2 | 9.4 |
| New Brunswick | 3.8 | 0.6 | 0.5 | 4.9 |
| Quebec | 2.5 | 0.6 | 0.4 | 3.5 |
| Ontario | 7.2 | 2.0 | 2.0 | 11.3 |
| Manitoba | 4.3 | 0.7 | 0.6 | 5.6 |
| Saskatchewan | 6.9 | 1.0 | 0.8 | 8.7 |
| Alberta | 3.6 | 0.7 | 0.6 | 4.9 |
| British Columbia | 4.4 | 1.2 | 0.8 | 6.4 |
| Yukon | 6.0 | 0.4 | 0.4 | 6.8 |
| Northwest Territories | 8.9 | 2.8 | 0.3 | 12.0 |
| Nunavut | 9.1 | 1.0 | 0.0 | 10.0 |

Table 5.1 indicates that 7.2% of households required some expenditure imputation (excluding the clothing section) at the national level, but nearly 75% of them had only one variable imputed. There are very few households that had more than one variable imputed (1.8%). At the provincial level, Quebec (3.5%), Prince Edward Island (4.1%) and New Brunswick (4.9%) have the lowest imputation rates. The highest rates are observed in Newfoundland and Labrador (11%), Ontario (11.3%) and the Northwest Territories (12.0%). The low percentage of persons for which some variables had to be imputed, combined with the low number of variables that had to be imputed when imputation was necessary, suggests that the imputed values should not have a strong impact on the estimates.

*5.1.2 Clothing Expenditure and Income Imputation by Province and Territory*

Since some respondents provide only totals for clothing expenditure and income variables, a two-step procedure is used to impute these variables (at the individual level). Individuals who only require imputation of certain components are imputed first. Then, they are followed by those for which totals are available but imputation on all components is required (see reference [1] for a more detailed description of this process).

The percentage of usable individuals (persons who are members of usable households) requiring imputation for an income variable are presented by province and territory in Table 5.2. The percentage of persons who had exactly one variable imputed, those who had two or more variables (but not all) imputed and the percentage of persons for which only total income was available (and hence required having all their components imputed) are shown. The total percentage of persons requiring some form of income imputation is also provided. The last column of Table 5.2 indicates this total percentage of persons requiring some form of imputation, but for the clothing expenditure variables.

**Table 5.2**
**Persons Requiring Income Imputation and Persons Requiring Clothing Expenditure Imputation by Province and Territory**

| Province | Percentage of persons requiring imputation for INCOME VARIABLES | | | | Percentage of persons requiring imputation for at least one of the 15 CLOTHING EXPENDITURE VARIABLES |
|---|---|---|---|---|---|
| | 1 income variable imputed | 2 or more income variables imputed (not all) | All income variables imputed (total income known) | **TOTAL** (any form of income imputation) | |
| **Canada** | **0.5** | **0.2** | **2.7** | **3.5** | **12.6** |
| Newfoundland and Labrador | 0.2 | 0.1 | 1.7 | 2.0 | 14.8 |
| Prince Edward Island | 0.0 | 0.1 | 5.1 | 5.2 | 13.3 |
| Nova Scotia | 0.5 | 0.3 | 3.1 | 3.9 | 15.2 |
| New Brunswick | 0.2 | 0.2 | 5.9 | 6.4 | 13.1 |
| Quebec | 0.3 | 0.1 | 1.4 | 1.8 | 9.8 |
| Ontario | 1.3 | 0.5 | 2.4 | 4.2 | 13.5 |
| Manitoba | 1.2 | 0.2 | 2.3 | 3.7 | 13.2 |
| Saskatchewan | 0.4 | 0.2 | 2.0 | 2.5 | 11.9 |
| Alberta | 0.2 | 0.3 | 3.2 | 3.6 | 15.1 |
| British Columbia | 0.6 | 0.3 | 2.9 | 3.8 | 9.8 |
| Yukon | 0.3 | 0.0 | 1.2 | 1.5 | 12.8 |
| Northwest Territories | 0.0 | 0.0 | 1.8 | 1.8 | 8.5 |
| Nunavut | 0.1 | 0.0 | 0.1 | 0.2 | 7.6 |

These results show that less than 4% of persons from usable households had some imputation performed on at least one income variable. For about 80% of them, the respondent gave the total income but all their components had to be imputed. For many of the remaining persons requiring imputation, only one component of income (one variable) had to be imputed. Provincially, the percentages of persons requiring some imputation on at least one income variable are also low, ranging from a low of 0.2% for Nunavut to a high of 6.4% for New Brunswick.

From the last column of the table, it can be seen that about 13% of persons required imputation for at least one of the clothing expenditure variables. The rates at the provincial and territorial levels range from 7.6% (for Nunavut) to around 15% (for Newfoundland and Labrador, Nova Scotia and Alberta). Almost all these people provided their total expenditure on clothing but required imputation of the components. The higher level of imputation required on clothing expenditure components implies that the estimates for these components could be greatly affected by imputation, while the effect on the estimates for total clothing expenditures will be negligible.

## References


[1] Tremblay, J. and Arsenault, S. 2001. *Methodology of the Survey of Household Spending*. Ottawa. Household Survey Methods Division, Statistics Canada.

[2] Wolter, K.M. 1985. *Introduction to Variance Estimation*. New York. Springer-Verlag.

[3] Statistics Canada, Income Statistics Division. 2001. *User Guide—Survey of Household Spending*. Catalogue no. 62F0026MIE2002002. Ottawa.

[4] Statistics Canada, Income Statistics Division. 2001. *Spending Patterns in Canada*. Catalogue no. 62-202-XPE. Ottawa.

[5] Statistics Canada, Household Survey Methods Division. 2001. *Labour Force Survey, Operations Report*. Survey 200112. Ottawa.

# Appendix A

# Algebraic notation

### 1. Nonresponse Adjustment

The subweight (i.e. the design weight adjusted for nonresponse) for a household k, denoted as $w_k^{NR}$, is

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{rate_g} \qquad with \qquad rate_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

where

$s_{g,r}$ is the set of respondents in nonresponse group g,

$s_{g,nr}$ is the set of nonrespondents (refusals, no contacts, unusable data) in nonresponse group g, and

• $\pi^{-1}$ is the design weight assigned to household k.

### 2. Calculation of the slippage rate

The slippage rate for a control group c, denoted as $rate_c$, is

$$rate_c = 100 * \frac{\left( \sum_{k \in s_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

where

$s_{c,r}$ is the set of respondents in control group c,

$w_k^{NR}$ is the subweight of household k, and

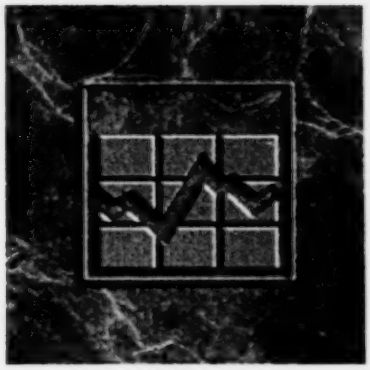
$t_c$ is the total of the auxiliary data for the control group c.

Nº 62F0026MIF au catalogue — Nº 001
ISSN: 1708-8887
ISBN: 0-662-75349-6

**Document de recherche**

**Série de documents de recherche sur les dépenses des ménages**

# Enquête sur les dépenses des ménages de 2001 - Indicateurs de qualité des données



2001

Division des méthodes d'enquêtes auprès des ménages
Immeuble R.H. Coats, Ottawa, K1A 0T6

Telephone: 613 951-7355

Statistique Canada Statistics Canada

Canada

## Comment obtenir d'autres renseignements

Veuillez communiquer avec les Services à la clientèle, Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355: revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

| | |
|---|---|
| **Service national de renseignements** | **1 800 263-1136** |
| **Service national d'appareils de télécommunications pour les malentendants** | **1 800 363-7629** |
| **Renseignements concernant le Programme des bibliothèques de dépôt** | **1 800 700-1033** |
| **Télécopieur pour le Progamme des bibliothèques de dépôt** | **1 800 889-9734** |
| **Renseignements par courriel** | **infostats@statcan.ca** |
| **Site Web** | **www.statcan.ca** |

## Renseignements sur les commandes et les abonnements

Le produit nº 62F0026MIF2003001 au catalogue est gratuit sur Internet. Les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/research_f.cgi.

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.

Statistique Canada
Division de la statistique du revenu

**Série de documents de recherche sur les dépenses des ménages**

# Enquête sur les dépenses des ménages de 2001 - Indicateurs de qualité des données

2001

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2003

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Novembre 2003

N° 62F0026MIF2003001 au catalogue

ISSN: 1708-8887

ISBN: 0-662-75349-6

Périodicité : Irrégulier

Ottawa

This publication is available in English upon request (Catalogue no. 62F0026MIE)

---

**Note de reconnaissance**

*Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.*

Ce document a été préparé par les membres de l'unité de l'Enquête sur les dépenses des ménages de la Division des méthodes d'enquêtes auprès des ménages (DMEM) :

Sylvie DeBlois
José Gaudet
Bruno Lapierre
Jessica Law
Christian Nadeau
Sylvain Nadon

Les auteurs remercient l'équipe responsable de l'Enquête sur les dépenses des ménages de la Division de la statistique du revenu (DSR) pour leur collaboration dans la préparation de ce document.

# Table des matières

**Faits saillants** ..... **5**

**Introduction** ..... **7**

**1. Les erreurs d'échantillonnage** ..... **8**
1.1 Les mesures de l'erreur d'échantillonnage ..... 8
1.2 Les coefficients de variation ..... 8
Tableau 1.1
Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen. ..... 9
Tableau 1.2
Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager ..... 10
1.3 Modèle pour dériver une approximation du CV ..... 10
1.4 Suppression des données non fiables dans les tableaux d'estimations ..... 12

**2. La non-réponse** ..... **12**
2.1 Les taux de non-réponse et les taux de vacance ..... 13
Tableau 2.1
Taux de non-réponse (%) et taux de vacance (%) par province et territoire ainsi qu'à l'échelle nationale ..... 14
2.2 La non-réponse selon le niveau d'urbanisation ..... 15
Tableau 2.2
Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation ..... 15
2.3 La non-réponse selon les strates de revenu ..... 16
Tableau 2.3
Comparaison des taux (%) de non-réponse et de vacance des strates à revenu élevé et à faible revenu par rapport aux autres strates ..... 16
2.4 L'ajustement pour la non-réponse ..... 16

**3. Erreurs de couverture** ..... **17**
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement ..... 18
Tableau 3.1
Taux de glissement nationaux selon le groupe d'âge-sexe ..... 19
Tableau 3.2
Taux de glissement pour les provinces et les territoires selon le groupe d'âge-sexe ..... 20
Tableau 3.3
Taux de glissement pour les provinces et les territoires selon la taille de ménage ..... 21
3.2 L'ajustement pour l'erreur de couverture de la population et du nombre de ménages ..... 21

**4. Les erreurs de réponse** ..... **23**

# Table des matières (fin)

**5. Les erreurs de traitement** ..... **24**
5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale ..... 26
Tableau 5.1
Ménages qui nécessitent l'imputation de dépenses par province et territoire ..... 27
Tableau 5.2
Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province et territoire ..... 28

**Références** ..... **29**

**Annexe A Notation algébrique** ..... **30**

# Faits saillants

## Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 1,9 % selon les provinces, et il est de 0,7 % à l'échelle nationale. Les CV sont plus élevés dans les territoires, soit 5,8 % au Yukon, 4,7 % dans les Territoires du Nord-Ouest et 8,1 % au Nunavut.

- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont inférieurs ou égaux à 2,4 % à l'échelle nationale, et généralement inférieurs à 6 % à l'échelle provinciale. Les résultats pour les différentes catégories de caractéristiques des logements et l'équipement ménager sont similaires. Les CV à l'échelle territoriale sont généralement plus élevés.

## La non-réponse

- Le taux de non-réponse est de 23,8 %. La non-réponse est due aux refus (15,2 %), aux ménages qu'on n'a pas réussi à contacter (5,2 %) et aux ménages dont les données ont été considérées inutilisables (3,4 %).

- Le taux de non-réponse final a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 19,3% en milieu rural, et un taux de 30,6% dans les centres urbains d'un million d'habitants et plus. Cette tendance s'observe également dans les taux de non-réponse à la collecte.

- L'analyse des taux de réponse finaux dans les strates d'aires géographiques à revenu élevé et à faible revenu provenant du plan d'échantillonnage révèle que le taux de non-réponse dans les strates à revenu élevé (37,5 %) est plus élevé que ceux observés dans les strates régulières (23,5 %) et dans les strates à faible revenu (18,6 %). Le taux de refus des strates à revenu élevé est environ le double de celui des strates à faible revenu.

## Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 6 % à l'échelle nationale. On observe aussi un sous-dénombrement des ménages pour chacune des provinces et territoires, les taux variant de 4,1 % à 21,6 %. C'est aux Territoires du Nord-Ouest et au Nunavut que le sous-dénombrement des ménages est le plus élevé.

- On observe un sous-dénombrement des personnes de 7,3 % à l'échelle nationale. On observe aussi un sous-dénombrement des personnes pour chacune des provinces et territoires, les taux variant de 4,7 % à 21,3 %. C'est aux Territoires du Nord-Ouest et au Nunavut que le sous-dénombrement des personnes est le plus élevé.

- Les taux de glissement nationaux pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il

y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 0,7 %, alors qu'il est de 9,2 % pour les adultes.

### Les erreurs de réponse

- Les erreurs de réponse comprennent, entre autres, les erreurs de rappel, l'erreur télescopique et les erreurs dues aux réponses par procuration. Dans l'EDM, le fardeau imposé au répondant, étant donné la longueur de l'entrevue, peut entraîner la fatigue du répondant et avoir un impact sur la qualité des données. Le temps total d'entrevue varie selon les caractéristiques des ménages et peut être de plus de cinq heures pour certains ménages. Le temps moyen de l'entrevue était une heure et dix minutes.

### Les erreurs de traitement reliées à l'imputation

#### i) Des variables de dépenses

- Il a fallu imputer au moins une variable de dépenses (excluant les dépenses vestimentaires) pour 7,2 % des ménages. Pour la majorité d'entre eux, seulement une ou deux variables parmi les 251 variables de dépenses ont été imputées.

- Environ 13 % des individus ont nécessité l'imputation de variables sur les dépenses vestimentaires mais pour la très grande majorité, le total était fourni par le répondant et seules les composantes ont été imputées.

#### ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour moins de 4 % des individus. Pour environ 80 % d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.

# Introduction

L'Enquête sur les dépenses des ménages (EDM) est une enquête annuelle qui collecte les dépenses et les revenus des ménages à l'aide d'entrevues personnelles. L'échantillon de l'EDM 2001 est composé de 25 740 logements[1] répartis dans les 10 provinces et les trois territoires. La collecte est effectuée au cours des mois de janvier, février et mars et les revenus et dépenses sont obtenus pour la période du 1er janvier au 31 décembre de l'année précédente. Suite à un remaniement en 1997, cette enquête remplace l'Enquête périodique sur les dépenses des familles et l'Enquête sur l'équipement ménager (avec ajustements aux questionnaires et aux échantillons).

L'EDM, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document vise donc à présenter les indicateurs de qualité produits pour l'enquête sur les dépenses des ménages de 2001. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tels que les coefficients de variation, les taux de non-réponse, les taux de glissement et les taux d'imputation.

Les indicateurs de qualité ont été classés selon les principaux types d'erreur que l'on retrouve dans une enquête. Les erreurs d'échantillonnage, c'est-à-dire les erreurs dues au fait que les inférences tirées au sujet de l'ensemble de la population sont fondées sur des renseignements recueillis auprès d'un échantillon de la population et non pas auprès de toute la population, sont présentées dans la section 1. Les sections suivantes couvrent les erreurs non dues à l'échantillonnage. La non-réponse et les erreurs de couverture sont d'abord discutées dans les sections 2 et 3. Les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [1].

[1] De ces 25 740 logements, on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 22 172 ménages auprès desquels on collecte les dépenses et les revenus.

# 1. Les erreurs d'échantillonnage

Les erreurs d'échantillonnage découlent du fait que les inférences que l'on tire au sujet de l'ensemble de la population à partir de l'enquête sont fondées sur l'information que l'on a recueillie auprès d'un échantillon de la population et non pas auprès de toute la population. Outre le plan de sondage et la méthode d'estimation qui ont été appliqués pour l'enquête sur les dépenses des ménages, la taille d'échantillon ainsi que la variabilité de chaque caractéristique sont des facteurs déterminants de l'erreur d'échantillonnage. Les caractéristiques qui sont rares ou qui sont distribuées de façon très différente dans la population auront une erreur d'échantillonnage plus grande que les caractéristiques qu'on observe plus fréquemment ou qui sont plus homogènes dans la population.

## 1.1 Les mesures de l'erreur d'échantillonnage

L'erreur-type est une mesure standard de l'erreur d'échantillonnage. L'erreur-type correspond au degré de variation de l'estimation considérant qu'un échantillon particulier a été choisi, plutôt qu'un autre, parmi tous les échantillons possibles de même taille sous le même plan de sondage. Comme l'EDM utilise un plan d'échantillonnage et une méthode d'estimation complexes, on estime l'erreur-type à partir d'une méthode de rééchantillonnage appelée le Jackknife. Pour plus de détails sur cette méthode, voir la référence [2].

Le coefficient de variation (CV) est également une mesure de fiabilité de l'estimation fréquemment utilisée. Il exprime tout simplement l'erreur-type en pourcentage de l'estimation. Ainsi, si on obtient une estimation Y pour une certaine caractéristique et que SE correspond à l'erreur-type estimée, le CV sera (SE/Y) x 100.

Finalement, l'erreur-type ou le coefficient de variation peuvent être utilisés pour dériver une autre mesure de précision des estimations, l'intervalle de confiance. Cette mesure indique le niveau de confiance selon lequel la valeur réelle dans la population pour une caractéristique qu'on a observée se trouve à l'intérieur de certaines limites. Un intervalle avec un niveau de confiance de 95 % correspond à l'estimation obtenue à partir de l'échantillon ± 2 erreur-type : (Y ± 2 SE).[2] Cela signifie que si on répétait l'échantillonnage un grand nombre de fois, chaque échantillon fournirait un intervalle différent et 95 % des intervalles contiendraient la vraie valeur de la caractéristique. De façon similaire, si on répétait l'échantillonnage, l'intervalle Y± SE contiendrait la vraie valeur dans 68 % des cas.

## 1.2 Les coefficients de variation

Des estimations des coefficients de variation sont calculées pour les estimations de plusieurs caractéristiques collectées dans l'EDM. Les CV pour les estimations de la moyenne des dépenses par ménage (pour les catégories de dépenses détaillées), ainsi que pour les caractéristiques des logements et l'équipement ménager, sont disponibles à l'échelle nationale, provinciale et territoriale dans la publication *Guide de l'utilisateur — Enquête sur les dépenses des ménages* (référence [3]).

[2] L'intervalle de confiance se calcule directement à partir du CV de façon similaire, soit Y ± 2 (CV x Y)/100.

On doit noter que les estimations de CV ne tiennent pas compte du fait que certaines données ont été imputées; par conséquent, les CV présentés peuvent sous-estimer les valeurs réelles. Pour la plupart des variables, le taux d'imputation est peu élevé (voir section 5) et les CV fournis représentent une bonne estimation des vrais CV. Cependant, il est important de tenir compte à la fois du CV et du taux d'imputation si on s'intéresse à la fiabilité de dépenses détaillées ayant un taux d'imputation élevé.

Le tableau 1.1 donne un aperçu des CV des estimations des moyennes par ménage à l'échelle provinciale et territoriale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories de dépenses sommaires et pour le revenu.

**Tableau 1.1**
**Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen**

| Catégories de dépenses sommaires | Can. | T.-N.-L. | Î.-P.-É. | N.-É. | N.-B. | Qc | Ont. | Man. | Sask. | Alb. | C.-B. | Yn | T. N.-O. | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dépenses totales | 0,7 | 1,6 | 1,9 | 1,2 | 1,2 | 1,2 | 1,4 | 1,2 | 1,4 | 1,4 | 1,2 | 5,7 | 4,6 | 8,1 |
| Dépenses de consommation | 0,6 | 1,4 | 1,7 | 1,1 | 1,3 | 1,1 | 1,3 | 1,4 | 1,4 | 1,5 | 1,2 | 5,0 | 3,7 | 6,9 |
| Alimentation | 0,5 | 1,1 | 1,6 | 1,1 | 1,1 | 0,9 | 1,1 | 1,2 | 1,3 | 1,2 | 1,2 | 3,7 | 2,8 | 4,7 |
| Logement | 0,8 | 1,8 | 3,0 | 1,6 | 2,4 | 1,5 | 1,4 | 2,0 | 1,9 | 2,1 | 1,5 | 5,8 | 6,8 | 12,6 |
| Entretien ménager | 0,9 | 2,3 | 4,1 | 1,8 | 2,2 | 1,8 | 1,8 | 2,2 | 2,2 | 2,0 | 1,8 | 6,3 | 6,9 | 7,8 |
| Ameublement | 2,4 | 3,5 | 5,5 | 3,4 | 4,1 | 3,7 | 4,7 | 3,9 | 3,6 | 5,8 | 4,4 | 9,1 | 5,2 | 10,5 |
| Habillement | 1,2 | 2,7 | 3,0 | 2,4 | 1,9 | 2,2 | 2,4 | 2,5 | 2,7 | 2,5 | 2,5 | 7,2 | 5,3 | 7,5 |
| Transport | 1,4 | 3,2 | 4,7 | 3,2 | 3,3 | 3,1 | 2,8 | 3,8 | 3,6 | 3,1 | 3,1 | 8,0 | 4,6 | 22,8 |
| Soins de santé | 1,3 | 3,1 | 6,4 | 2,6 | 3,0 | 2,1 | 3,5 | 2,7 | 2,9 | 2,2 | 2,7 | 10,4 | 10,6 | 31,3 |
| Soins personnels | 0,9 | 2,0 | 2,5 | 2,0 | 1,8 | 2,0 | 1,8 | 1,9 | 2,0 | 2,1 | 2,0 | 5,1 | 4,0 | 8,0 |
| Loisirs | 1,8 | 3,3 | 3,9 | 3,3 | 3,4 | 3,1 | 3,6 | 4,1 | 3,3 | 4,4 | 4,3 | 7,7 | 6,8 | 9,6 |
| Lecture et autres imprimés | 1,6 | 3,8 | 4,8 | 3,6 | 3,1 | 3,1 | 3,3 | 3,4 | 3,2 | 3,3 | 3,1 | 12,6 | 9,1 | 9,3 |
| Éducation | 3,1 | 8,3 | 10,9 | 10,8 | 8,0 | 5,5 | 5,6 | 8,5 | 7,9 | 5,7 | 6,4 | 19,4 | 24,9 | 36,3 |
| Tabac, boissons alcoolisées | 1,5 | 4,3 | 5,6 | 3,9 | 5,1 | 3,0 | 3,0 | 4,3 | 4,5 | 3,7 | 4,1 | 6,5 | 5,8 | 8,3 |
| Jeux de hasard (net) | 3,4 | 10,4 | 12,9 | 8,4 | 7,6 | 6,8 | 6,5 | 10,7 | 13,6 | 9,8 | 8,6 | 22,4 | 16,6 | 17,6 |
| Dépenses diverses | 3,0 | 6,9 | 8,2 | 6,4 | 7,5 | 6,2 | 6,2 | 6,1 | 5,2 | 5,1 | 6,2 | 18,9 | 10,4 | 34,0 |
| Impôts personnels | 1,7 | 4,3 | 4,4 | 2,6 | 2,1 | 2,2 | 3,5 | 2,5 | 2,7 | 3,6 | 2,8 | 9,2 | 8,3 | 13,7 |
| Assurances personnelles | 1,5 | 2,4 | 3,1 | 2,5 | 1,7 | 2,7 | 3,1 | 1,9 | 2,4 | 2,6 | 2,5 | 12,5 | 6,1 | 8,5 |
| Dons et contributions | 3,6 | 6,9 | 9,7 | 6,9 | 10,0 | 8,7 | 6,1 | 5,9 | 6,9 | 7,9 | 10,0 | 13,3 | 17,4 | 27,4 |
| Revenu | 0,7 | 1,4 | 1,7 | 1,0 | 0,8 | 1,3 | 1,4 | 0,9 | 1,1 | 1,8 | 1,0 | 8,5 | 6,8 | 11,4 |

Les coefficients de variation de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 1,9 % selon les provinces et il est de 0,7 % à l'échelle nationale. Les CV sont plus élevés dans les territoires, soit 5,8 % au Yukon, 4,7 % dans les Territoires du Nord-Ouest et 8,1 % au Nunavut.

Pour les différentes catégories de dépenses sommaires, les CV à l'échelle nationale sont inférieurs ou égaux à 2,4 % à l'exception des dépenses dans les catégories suivantes: *éducation, jeux de hasard, dépenses diverses,* ainsi que *dons en argent et contributions.* Ces catégories de dépenses représentent respectivement 1,6 %, 0,5 %,

1,5 % et 2,2 % du total des dépenses (données non présentées). De plus, si on fait exception de ces catégories, les CV sont généralement inférieurs ou égaux à 6 % à l'échelle provinciale.

Le tableau 1.2 donne un aperçu des CV pour les caractéristiques des logements et l'équipement ménager à l'échelle provinciale et territoriale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories.

**Tableau 1.2**
**Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager**

| CATÉGORIES | Can. | T.-N.-L. | Î.-P.-É. | N.-É. | N.-B. | Qc | Ont. | Man. | Sask. | Alb. | C.-B. | Yn | T. N.-O. | Nt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Propriétaire | 0,9 | 1,6 | 3,1 | 1,7 | 1,7 | 1,9 | 1,8 | 1,9 | 1,8 | 1,8 | 1,6 | 8,6 | 13,9 | 16,0 |
| Locataire | 1,6 | 5,1 | 7,2 | 4,5 | 4,5 | 2,7 | 3,3 | 4,7 | 4,6 | 4,5 | 2,9 | 15,0 | 13,5 | 4,1 |
| Machine à laver | 0,6 | 0,9 | 2,0 | 1,3 | 1,1 | 1,0 | 1,5 | 1,7 | 1,2 | 1,3 | 1,4 | 3,0 | 4,3 | 5,4 |
| Sécheuse | 0,7 | 1,1 | 2,2 | 1,4 | 1,3 | 1,1 | 1,5 | 1,7 | 1,1 | 1,4 | 1,3 | 3,5 | 4,2 | 5,7 |
| Lave-vaisselle | 1,1 | 4,3 | 5,1 | 3,1 | 3,2 | 2,1 | 2,5 | 2,8 | 2,9 | 2,2 | 2,0 | 11,9 | 13,1 | 33,6 |
| Congélateur | 1,1 | 1,4 | 2,5 | 1,9 | 2,0 | 2,5 | 2,3 | 1,7 | 1,7 | 1,8 | 2,1 | 5,8 | 5,1 | 9,4 |
| Four à micro-ondes | 0,4 | 0,9 | 1,2 | 0,7 | 0,7 | 0,9 | 0,7 | 0,7 | 0,9 | 0,6 | 0,9 | 2,9 | 2,9 | 6,7 |
| Téléphone cellulaire | 1,2 | 3,9 | 5,5 | 3,4 | 4,0 | 2,9 | 2,1 | 2,9 | 2,9 | 2,1 | 2,5 | 17,2 | 15,1 | 24,4 |
| Lecteur de disques compacts | 0,7 | 1,7 | 3,3 | 1,6 | 1,6 | 1,6 | 1,4 | 1,6 | 1,5 | 1,1 | 1,3 | 4,3 | 3,1 | 3,5 |
| Télédistribution | 1,0 | 2,1 | 5,3 | 2,3 | 2,4 | 2,4 | 2,0 | 2,3 | 2,8 | 1,8 | 1,6 | 10,3 | 11,0 | 3,5 |
| Ordinateur personnel | 0,9 | 2,6 | 4,3 | 2,3 | 2,7 | 2,2 | 1,7 | 2,3 | 2,6 | 1,7 | 1,7 | 8,3 | 5,5 | 14,0 |
| Connexion tél. régulière à un ordinateur (Modem) | 1,9 | 5,1 | 7,0 | 4,6 | 4,0 | 3,8 | 3,4 | 5,4 | 5,6 | 4,9 | 4,4 | 12,4 | 11,3 | 19,2 |
| Utilisation d'Internet (maison) | 1,1 | 3,2 | 5,4 | 2,8 | 3,4 | 2,5 | 2,0 | 2,6 | 3,3 | 2,2 | 2,2 | 8,5 | 5,4 | 17,7 |
| Véhicules possédés (un) | 1,5 | 3,1 | 4,7 | 3,2 | 3,4 | 2,9 | 3,2 | 3,2 | 3,1 | 3,2 | 2,6 | 8,3 | 6,3 | 14,7 |
| Véhicules possédés (2 ou plus) | 1,4 | 4,2 | 4,6 | 3,4 | 3,7 | 3,3 | 3,0 | 3,2 | 2,6 | 2,4 | 2,5 | 9,1 | 16,4 | 37,2 |

Les coefficients de variation pour les caractéristiques des logements et l'équipement ménager sont généralement inférieurs à 5 % à l'échelle provinciale, avec quelques exceptions dans la catégorie *Connexion téléphonique régulière à un ordinateur (Modem)*. Les CV sont plus élevés dans les territoires où l'on retrouve une proportion moindre des équipements.

Les CV pour les caractéristiques des logements et l'équipement ménager à l'échelle nationale sont inférieurs ou égaux à 1,9 %.

## 1.3 Modèle pour dériver une approximation du CV

Des estimations pour différents domaines d'intérêt (par ex., par quintile de revenu) sont disponibles pour les catégories de dépenses sommaires dans la publication *Les habitudes de dépenses au Canada* (référence [4]). Des estimations pour différents domaines d'intérêt pour les catégories de dépenses détaillées sont également disponibles sur demande auprès de la Division de la statistique du revenu. (Pour plus de

détails sur les tableaux disponibles auprès de la Division de la statistique du revenu, voir les références [3] ou [4].) Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui pourraient intéresser les utilisateurs.

*1.3.1 Approximation du CV pour les estimations des domaines*

Il est toutefois possible de calculer une approximation du CV en utilisant une relation entre le nombre de ménages de l'échantillon qui ont déclaré des dépenses pour une catégorie et le CV à un niveau agrégé. Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de ménages déclarant une dépense, est illustrée ci-après.

***Formule d'approximation du CV pour un domaine (un sous-groupe de la population)***

Si CV(Y) représente le CV pour l'estimation de la moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un quintile de revenu, un niveau d'urbanisation, ...) à partir de l'équation suivante :

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

Où

$n$ : *le nombre de ménages dans l'échantillon*
$P$ : *l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans la population*
$n_d$ : *le nombre de ménages de l'échantillon dans le domaine d*
$P_d$ : *l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans le domaine d*

Généralement on utilise le CV, la taille $n$ et la proportion $P$ à l'échelle nationale pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule province (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle provinciale puisque les CV provinciaux sont publiés pour l'EDM de 2001 (référence [3]). Il est important de se rappeler que la valeur du CV obtenue par cette approche n'est qu'une approximation du CV.

*1.3.2 Approximation du CV à partir du fichier de microdonnées*

Les utilisateurs du fichier de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV des estimations, qui sera généralement plus performante que celle présentée dans la section précédente pour les CV des catégories de dépenses détaillées. Cette approche est décrite en détails dans la documentation accompagnant le fichier de microdonnées de 2001. Elle est utilisable seulement à partir

du fichier de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

Le document sur la qualité des données de l'enquête de 1997 renferme des résultats de l'évaluation de la performance de ces deux méthodes d'approximation du CV.

### 1.4 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, on aimerait l'utiliser pour déterminer si les estimations devraient être publiées ou non. On considère que les estimations dont le CV est estimé à plus de 33 % ne sont pas suffisamment fiables pour être publiées. Toutefois les estimations des CV ne sont pas calculées pour une grande partie des estimations publiées, la règle de suppression pour les estimations de dépenses est donc basée sur le nombre de ménages qui déclarent une valeur supérieure à zéro.[3]

On peut démontrer que les CV sont en général inférieurs à 33 % si le nombre de ménages qui déclarent une dépense est supérieur à 30. Comme il s'agit d'une règle approximative, certaines estimations peuvent être publiées même si le CV est supérieur à 33 % et certaines estimations ne seront pas publiées malgré un CV inférieur à 33 %. Le document sur la qualité des données de l'EDM 1997 donne les résultats de l'évaluation des risques d'erreurs de l'utilisation de la règle de suppression.

## 2. La non-réponse

Les erreurs dues à la non-réponse découlent du fait que certains répondants potentiels ne fournissent pas l'information nécessaire ou que cette information s'avère inutilisable. Lorsque le répondant a omis de répondre à certaines questions seulement, on parle alors de non-réponse partielle. Dans ce cas, les données manquantes sont imputées. Les erreurs liées à l'imputation sont présentées dans la section 5 portant sur les erreurs de traitement. Dans la présente section, la non-réponse englobe la non-réponse à la collecte, due principalement à l'impossibilité de contacter le ménage ou au refus des membres du ménage de participer à l'enquête, que ce soit partiellement ou complètement, ainsi que les ménages pour lesquels les données ont été collectées mais sont inutilisables.

Le principal impact de la non-réponse sur la qualité des données est qu'elle peut induire un biais dans les estimations si les caractéristiques des répondants et des non-répondants diffèrent et que cette différence a un impact sur les caractéristiques étudiées. Les taux de non-réponse peuvent être calculés facilement mais ils n'ont qu'une valeur indicative à l'égard de la qualité des données car ils ne permettent pas de mesurer l'importance du biais associé aux estimations. L'ampleur de la non-réponse peut être considérée comme une évaluation des risques de biais dans les estimations.

[3] En pratique, on utilise l'estimation de la proportion des ménages déclarant une dépense que l'on multiplie par la taille d'échantillon.

## 2.1 Les taux de non-réponse et les taux de vacance

Dans l'EDM, comme les unités sélectionnées sont des logements, les intervieweurs doivent d'abord identifier les logements inadmissibles, c'est-à-dire les logements occupés par des personnes ne faisant pas partie de la population cible, les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise) et les logements vacants (inoccupés, saisonniers ou en construction).

Parmi les logements admissibles, on évalue ensuite la proportion des ménages qui n'ont pas répondu à l'enquête, qu'on appelle le taux de non-réponse à la collecte. Ces derniers comprennent les ménages qui ont refusé de participer à l'enquête et les ménages où il a été impossible d'établir un contact avec les répondants parce qu'ils étaient absents ou encore à cause de circonstances spéciales (problème de langue, maladie ou décès).

Toujours parmi les logements admissibles, on détermine également le taux de données inutilisables. Les données inutilisables correspondent au nombre de ménages pour lesquels le questionnaire était au moins partiellement complet mais qui ont été rejetés lors du traitement des données. Il existe deux causes principales de rejet. D'abord lorsqu'une partie importante des questions sur le revenu ou des questions sur les dépenses a été laissée sans réponse, le questionnaire est classé incomplet et n'est pas utilisé. L'autre source de rejet correspond aux questionnaires pour lesquels la différence entre les entrées (revenu et autres sources d'argent du ménage) et les déboursés (dépenses et variation nette des actifs et passifs) est supérieure à 20 %. Ces questionnaires sont également exclus de l'estimation et considérés comme de la non-réponse.

Le tableau 2.1 présente le taux de non-réponse final, le taux de non-réponse à la collecte, ventilé selon les refus et les non-contacts ainsi que le taux de données inutilisables, ventilé selon les incomplets et les non balancés. On y trouve également le taux de vacance. Ces taux sont fournis tant à l'échelle nationale que provinciale et territoriale.

**Tableau 2.1**
**Taux de non-réponse (%) et taux de vacance (%) par province et territoire ainsi qu'à l'échelle nationale**

| Province ou Territoire | Taux de vacance | Taux de non-réponse À la collecte | | | Taux de données inutilisables | | | Taux de non-réponse final (à l'estimation) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | Pas de contact | Refus | TOTAL | Incom-plet | Non balancé | |
| **Canada** | 11,5 | 20,4 | 5,2 | 15,2 | 3,4 | 1,9 | 1,5 | 23,8 |
| Terre-Neuve-et-Labrador | 18,4 | 16,1 | 5,4 | 10,8 | 2,1 | 1,1 | 1,0 | 18,3 |
| Île-du-Prince-Édouard | 19,3 | 15,3 | 4,6 | 10,7 | 0,8 | 0,4 | 0,4 | 16,0 |
| Nouvelle-Écosse | 13,5 | 21,6 | 6,7 | 14,9 | 4,3 | 3,0 | 1,3 | 25,9 |
| Nouveau-Brunswick | 15,9 | 18,4 | 4,5 | 14,0 | 2,8 | 1,4 | 1,4 | 21,2 |
| Québec | 6,9 | 21,1 | 6,5 | 14,6 | 1,5 | 0,9 | 0,6 | 22,7 |
| Ontario | 7,7 | 28,1 | 7,3 | 20,8 | 7,7 | 4,8 | 2,9 | 35,8 |
| Manitoba | 13,8 | 15,4 | 3,3 | 12,1 | 1,8 | 0,6 | 1,2 | 17,2 |
| Saskatchewan | 10,8 | 17,5 | 4,1 | 13,4 | 2,9 | 1,2 | 1,7 | 20,4 |
| Alberta | 5,3 | 19,3 | 3,2 | 16,1 | 3,5 | 1,5 | 2,0 | 22,8 |
| Colombie-Britannique | 7,7 | 22,3 | 3,8 | 18,5 | 3,1 | 1,8 | 1,3 | 25,4 |
| Yukon | 31,4 | 24,1 | 3,8 | 20,3 | 3,2 | 2,3 | 0,9 | 27,2 |
| Territoires du Nord-Ouest | 13,3 | 11,7 | 6,1 | 5,6 | 1,5 | 0,7 | 0,7 | 13,1 |
| Nunavut | 12,6 | 15,0 | 7,7 | 7,3 | 0,4 | 0,4 | 0,0 | 15,4 |

Le taux de non-réponse final au Canada est de 23,8 %. Il est dû aux refus (15,2 %), aux ménages qu'il a été impossible de rejoindre (5,2 %), et finalement aux ménages dont les données étaient inutilisables (3,4 %). Pour presque toutes les provinces, les refus représentent la cause majeure de non-réponse, suivis des non-contacts, puis des données inutilisables. Les seules exceptions sont l'Ontario et l'Alberta pour lesquelles la proportion de données inutilisables est légèrement supérieure à la proportion de non-contacts.

Les taux de non-réponse finaux varient selon les provinces. C'est aux Territoires du Nord-Ouest que l'on observe le taux de non-réponse le plus bas, soit de 13,1 %. C'est également pour ce territoire que la proportion de refus est la plus faible (5,6 %). Les taux de non-réponse de Terre-Neuve-et-Labrador, de l'Île-du-Prince-Édouard, du Manitoba et du Nunavut sont inférieurs à 20 %, tandis que les taux sont supérieurs à 20 % pour les autres provinces et territoires. Le taux de non-réponse de l'Ontario est particulièrement élevé atteignant presque 36 %. C'est aussi dans cette province que l'on trouve le plus haut taux de refus (20,8 %).

Le taux de vacance est présenté au tableau 2.1, mais on doit considérer que les logements vacants ne contribuent pas au biais de l'échantillon dans la mesure où ils sont identifiés correctement. En analysant les taux de vacance, on décèle les problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDM de 2001 est de 11,5 %. Le taux de vacance de l'EDM est légèrement plus bas que celui de l'Enquête sur la population active (EPA) pour la même période. Notons que le taux de vacance plus élevé (31,4 %) au Yukon est dû dans une certaine mesure à une répartition de l'échantillon différente en 2001 pour ce territoire et dans une moindre mesure aux changements apportés par l'EPA quant à la couverture au Yukon.

## 2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux à l'échelle nationale sont présentés par niveau d'urbanisation dans le tableau 2.2.[4]

**Tableau 2.2**
**Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation**

| Catégorie d'urbanisation | Taux de vacance | Taux de non-réponse à la collecte | | | Taux de données inutilisables | | | Taux de non-réponse final (à l'estimation) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | Pas de contact | Refus | TOTAL | Incom-plet | Non balancé | |
| **URBAIN** | | | | | | | | |
| 1 000 000 ou plus | 4,1 | 26,8 | 7,8 | 19,0 | 3,7 | 2,1 | 1,6 | 30,6 |
| 500 000 à 999 999 | 3,9 | 18,7 | 3,3 | 15,3 | 2,7 | 1,0 | 1,7 | 21,4 |
| 250 000 à 499 999 | 6,3 | 26,0 | 6,9 | 19,2 | 5,9 | 3,8 | 2,2 | 32,0 |
| 100 000 à 249 999 | 7,9 | 22,0 | 5,0 | 17,0 | 3,3 | 2,1 | 1,1 | 25,3 |
| 30 000 à 99 999 | 5,3 | 19,1 | 3,7 | 15,4 | 3,3 | 2,1 | 1,1 | 22,3 |
| Moins de 30 000 | 12,5 | 17,1 | 4,6 | 12,5 | 3,0 | 1,6 | 1,4 | 20,1 |
| **RURAL** | 23,4 | 16,6 | 4,6 | 12,0 | 3,2 | 1,8 | 1,4 | 19,8 |
| **TOTAL** | 11,5 | 20,4 | 5,2 | 15,2 | 3,4 | 1,9 | 1,5 | 23,8 |

Le taux de non-réponse final augmente généralement avec le niveau d'urbanisation. Selon les résultats du tableau 2.2, il n'y a que le groupe des 500 000 à 999 999 et, dans une moindre mesure, le groupe des 1 000 000 ou plus, qui ne suivent pas cette tendance.

Le taux de non-réponse à la collecte a lui aussi tendance à croître avec le niveau d'urbanisation. Il y a un écart important de près de 10% entre les catégories d'urbanisation «Moins de 30 000» et «1 000 000 ou plus». Les refus comptent pour plus de 60 % de la non-réponse totale à tous les niveaux d'urbanisation. Bien que la catégorie d'urbanisation « 500 000 à 999 999 » possède le plus faible taux de non-réponse totale pour les régions urbaines de 30 000 habitants ou plus, c'est pour celle-ci que la proportion de non-réponse due au refus des ménages de participer à l'enquête est la plus élevée (71,5 %).

En examinant les taux de vacance par niveau d'urbanisation, il ressort que le taux de vacance est près de deux fois plus élevé en région rurale (23,4 %) que dans les régions urbaines à faible population (12,5 %). Ces dernières régions ont aussi un taux de vacance supérieur à celui des régions urbaines à plus forte population. Ce phénomène s'observe également dans l'EPA et s'explique sans doute par un plus grand nombre de logements saisonniers en milieu rural. Comme l'échantillon de l'EDM est plus concentré dans les régions urbaines à forte population que celui de l'EPA, on peut s'attendre à observer un taux de vacance national légèrement inférieur.

[4] Des tableaux portant sur les taux de non-réponse selon le niveau d'urbanisation par province sont disponibles sur demande auprès de la Division des méthodes d'enquêtes auprès des ménages.

## 2.3 La non-réponse selon les strates de revenu

Il est impossible de comparer le taux de réponse selon le revenu car cette information n'est pas accessible pour les non-répondants. Toutefois, le plan d'échantillonnage de l'EPA, utilisé pour l'EDM, a été conçu de sorte à former, dans neuf grandes villes, des strates d'aires géographiques où le revenu moyen des ménages excède 100 000 $, et dans sept grandes villes, des strates composées d'appartements habités par des ménages ayant un revenu moyen inférieur à 20 000 $. Même si le nombre de telles strates est petit et qu'il compte pour seulement un petit nombre de logements dans l'échantillon de l'EDM (environ 580 pour les strates de revenus élevés et 160 pour les strates de revenus faibles, soit moins de 3 % de l'échantillon), la comparaison des taux de non-réponse dans ces deux groupes par rapport à l'ensemble des autres strates est révélatrice. Le tableau 2.3 présente ces résultats.

**Tableau 2.3**
**Comparaison des taux (%) de non-réponse et de vacance des strates à revenu élevé et à faible revenu par rapport aux autres strates**

| Type de strate en fonction du revenu | Taux de vacance | Taux de non-réponse à la collecte | | | Taux de données inutilisables | | | Taux de non-réponse final (à l'estimation) |
|---|---|---|---|---|---|---|---|---|
| | | TOTAL | Pas de contact | Refus | TOTAL | Incom-plet | Non balancé | |
| Revenu élevé | 2,9 | 31,5 | 8,8 | 22,7 | 5,9 | 1,8 | 4,1 | 37,5 |
| Régulières | 11,7 | 20,1 | 5,0 | 15,1 | 3,3 | 1,9 | 1,4 | 23,5 |
| Revenu faible | 5,1 | 17,9 | 6,2 | 11,7 | 0,7 | 0,7 | 0,0 | 18,6 |
| **TOTAL** | 11,5 | 20,4 | 5,2 | 15,2 | 3,4 | 1,9 | 1,5 | 23,8 |

Dans les strates à revenu élevé, le taux de non-réponse final (37,5 %) est de beaucoup supérieur à celui dans les strates régulières. Le taux de refus des strates à revenu élevé est de plus de 20%, ce qui est environ le double de celui des strates à faible revenu.

Les ménages des strates à faible revenu ont un taux de non-réponse à la collecte similaire à celui des strates régulières. Toutefois, les strates régulières ont un taux de non-réponse final de 26% supérieur à celui des strates à faible revenu. Ceci s'explique en grande partie par le fait qu'il n'y a presque pas de données inutilisables pour les strates à faible revenu.

Comme pour les enquêtes de 1997 à 2000, le taux de vacance est plus élevé pour les strates régulières que pour chacun des deux autres types de strate. Dans les strates à revenu élevé et à revenu faible, les taux de vacance sont semblables.

## 2.4 L'ajustement pour la non-réponse

Pour compenser la non-réponse, les poids de l'EDM sont gonflés par l'inverse du taux de réponse pondéré à l'intérieur de certains groupes définis en fonction des différents niveaux d'urbanisation dans chaque province. Les taux pondérés diffèrent des taux présentés dans cette section puisqu'ils tiennent compte du poids de sondage de chaque ménage. Une description algébrique de l'ajustement pour la non-réponse est présentée à l'annexe A.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par niveau d'urbanisation mentionnées dans la section 2.2. Il permettra de réduire le biais dans la mesure où les caractéristiques des répondants et des non-répondants sont similaires pour un même niveau d'urbanisation.

## 3. Erreurs de couverture

Lors de la conception de l'enquête, on a défini la population visée, qu'on appelle la population cible. Rappelons d'abord la définition de cette population cible pour l'EDM puisqu'une bonne compréhension de la population cible est nécessaire à une bonne interprétation des données de l'enquête. Il est important de préciser que l'EDM utilise la base de sondage de l'enquête sur la population active (EPA).

*La population cible*

> La population cible correspond aux individus vivant dans les ménages privés. Les pensionnaires d'établissement institutionnel tel que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont donc exclus. On exclut aussi les représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques. L'enquête couvre donc près de 98 % de la population dans les 10 provinces. Au Yukon, la couverture de la population est passé de 81 % en 1999 à 88 % en 2001. Cette hausse est due aux modifications apportées à la couverture de ce territoire pour l'EPA. Ainsi, certaines communautés et régions non organisées qui étaient exclues sont dorénavant incluses. La couverture des Territoires du Nord-Ouest correspond à 92 % de la population et la couverture du Nunavut est de 89 %.[5]

Les erreurs de couverture proviennent d'une représentation inadéquate de la population cible à partir des unités de la base de sondage. Certaines unités de la population cible peuvent être omises de la base de sondage, on parle alors d'un sous-dénombrement. D'autres unités qui ne sont pas dans la population cible peuvent être incluses par erreur ou certaines unités peuvent être incluses plus d'une fois, ces unités sont responsables du surdénombrement.

[5] Pour ce qui est de la proportion de ménages, la couverture du Yukon, des Territoires du Nord-Ouest et du Nunavut correspondent respectivement à 88 %, 93 % et 90 % des ménages.

## 3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement

Dans l'EDM, la sélection de l'échantillon est effectuée en utilisant une liste de logements dans chaque grappe sélectionnée. L'omission de logements lors de la création de la liste, les nouveaux logements qui s'ajoutent entre la création de la liste et la visite des intervieweurs (principalement dans les secteurs en développement) ainsi que la classification erronée de logements vacants contribuent au sous-dénombrement. Le fait d'inclure des logements qui ne seraient pas dans les limites de la grappe est une source de surdénombrement. De façon similaire, des erreurs peuvent se glisser lors de la collecte des données, lors de l'identification des personnes qui sont membres du ménage sélectionné. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

Une bonne représentation de la population cible est indispensable à la production d'estimations de dépenses réalistes. Le nombre de personnes par ménage est également une caractéristique importante dans l'estimation des dépenses moyennes des ménages. Il est donc nécessaire que l'échantillon, non seulement représente adéquatement les individus de la population cible, mais également la distribution des ménages selon leur taille.

En 1999, une nouvelle méthode de pondération a été introduite utilisant de nouveaux groupes de contrôle. Cette nouvelle méthode nous permet de mieux corriger la représentativité de la population visée en ciblant des groupes d'âge et de sexe mieux délimités que ceux utilisés précédemment et pour lesquels la couverture varie d'un groupe à l'autre.

On observe généralement un sous-dénombrement net du nombre de personnes et du nombre de ménages dans l'EDM, que l'on rectifie par un ajustement des poids en utilisant des données auxiliaires, qui sont basées sur des estimations démographiques post-censitaires. Le taux de glissement (voir annexe A) est une mesure du pourcentage d'écart entre les estimations provenant de ces données auxiliaires et les estimations de l'enquête calculées en utilisant des poids non ajustés avec ces données.[6] Les taux de glissement par groupe d'âge-sexe à l'échelle nationale, provinciale et territoriale sont présentés dans les tableaux 3.1 et 3.2 alors que les taux de glissement selon les tailles de ménages que l'on utilise lors de l'ajustement des poids se trouvent dans le tableau 3.3. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête.

[6] On utilise les sous-poids, c'est-à-dire les poids de sondage ajustés pour la non-réponse (voir Annexe A).

**Tableau 3.1**
**Taux de glissement nationaux selon le groupe d'âge-sexe**

| Taux de glissement national par groupe d'âge-sexe | | | | |
|---|---|---|---|---|
| | Âge | Sexe | | Total |
| | | Hommes | Femmes | |
| Canada[1] | 0-6 ans | 0,5 | 5,2 | 2,8 |
| | 7-17 ans | -3,5 | -1,6 | -2,6 |
| | 18-24 ans | -6,5 | -3,2 | -4,9 |
| | 25-34 ans | -13,4 | -10,2 | -11,8 |
| | 35-54 ans | -13,6 | -7,9 | -10,7 |
| | 55-59 ans | -7,3 | -15,9 | -11,7 |
| | 60-64 ans | -12,4 | -4,7 | -8,4 |
| | 65-69 ans | -6,4 | -0,8 | -3,5 |
| | 70 ans et + | -9,4 | -2,5 | -5,4 |
| | **Total** | **-9,2** | **-5,4** | **-7,3** |

[1] Exclut les territoires.

Pour l'EDM de 2001, le taux de sous-dénombrement de la population était 7,3 %. En analysant le tableau 3.1 au niveau des groupes d'âge, on constate qu'à l'échelle nationale, les taux de glissement pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 0,7 %, alors qu'il est de 9,2 % pour les adultes (données non présentées).

Les taux les plus élevés à l'échelle nationale sont observés chez les femmes de 55 à 59 ans, ainsi que chez les hommes de 25 à 34 ans et de 35 à 54 ans. À l'exception de la catégorie des 55 à 59 ans, le taux de sous-dénombrement chez les femmes est toujours inférieur à celui des hommes.

Tel que mentionné précédemment, l'EDM utilise la base de sondage de l'enquête sur la population active (EPA). Pour une même période, le taux de sous-dénombrement de l'EPA à l'échelle nationale est de 9,8 % (référence [5]), ce qui est légèrement supérieur au taux de 8,8 % de l'EDM observé lorsqu'on se restreint aux personnes de 15 ans et plus (données non présentées).

**Tableau 3.2**
**Taux de glissement pour les provinces et les territoires selon le groupe d'âge-sexe**

| Taux de glissement par groupe d'âge-sexe | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexe | Âge | Terre-Neuve-et-Labrador | Île-du-Prince-Édouard | Nouvelle-Écosse | Nouveau-Brunswick | Québec | Ontario | Manitoba | Saskatchewan | Alberta | Colombie-Britannique | Yukon | Territoires du Nord-Ouest | Nunavut |
| Homme | 0-6 | -0,1 | -21,1 | -2,0 | -8,1 | 12,0 | 0,2 | -14,8 | 6,4 | -8,3 | -4,7 | 7,4 | -18,3 | -19,4 |
| | 7-17 | -17,4 | 5,1 | -3,3 | -7,4 | -10,9 | 7,3 | -6,8 | 1,6 | -8,1 | -18,6 | | | |
| | 18-24 | -27,1 | -31,6 | -18,0 | -23,0 | -1,5 | 1,9 | -13,9 | -31,6 | -21,9 | -9,0 | | | |
| | 25-34 | -43,3 | -13,8 | -23,8 | -25,9 | -7,8 | -9,6 | -17,2 | -12,7 | -13,7 | -25,6 | | | |
| | 35-54 | -14,0 | -15,5 | -11,9 | -15,5 | -12,5 | -11,9 | -11,3 | -17,5 | -13,7 | -19,9 | | | |
| | 55-59 | -18,6 | -11,3 | -15,9 | -6,0 | -1,3 | -8,8 | -11,8 | -24,3 | -1,5 | -10,2 | -18,1 | -27,4 | -20,3 |
| | 60-64 | -0,3 | -6,5 | -12,1 | -38,1 | -7,7 | -15,7 | -16,4 | -28,5 | -11,5 | -4,8 | | | |
| | 65-69 | 15,5 | -15,5 | 15,2 | 4,5 | -5,4 | -11,7 | -10,8 | -13,1 | 2,5 | -5,1 | | | |
| | 70 + | 9,2 | -3,9 | -11,6 | -4,6 | -12,7 | -12,7 | -0,3 | 2,0 | -1,9 | -7,3 | | | |
| | **Total** | **-16,1** | **-12,6** | **-11,2** | **-14,8** | **-7,5** | **-6,3** | **-11,3** | **-11,9** | **-11,4** | **-15,5** | **-11,5** | **-24,6** | **-19,9** |
| Femme | 0-6 | -21,7 | 39,3 | -1,6 | 0,8 | 14,5 | 3,8 | 7,9 | 8,6 | 4,9 | -4,3 | | | |
| | 7-17 | -9,9 | -1,9 | 0,4 | -4,5 | -4,2 | 8,3 | -10,4 | -2,6 | -10,8 | -16,2 | 10,3 | -15,8 | -22,8 |
| | 18-24 | -30,4 | -34,8 | -18,5 | -16,3 | 14,7 | -5,2 | -14,9 | -24,6 | 5,6 | -17,7 | | | |
| | 25-34 | -27,3 | -8,1 | -15,9 | -11,8 | -2,1 | -11,7 | -2,0 | 0,4 | -11,9 | -18,4 | | | |
| | 35-54 | -11,6 | -12,0 | -6,3 | -5,9 | -9,9 | -3,0 | -7,5 | -7,4 | -8,7 | -17,8 | | | |
| | 55-59 | -8,0 | 4,5 | -10,1 | -10,5 | -2,1 | -23,3 | -27,6 | -28,2 | -17,6 | -20,7 | -10,9 | -19,0 | -13,3 |
| | 60-64 | 12,4 | -9,9 | 15,9 | 2,2 | -7,0 | -7,3 | -12,8 | -5,6 | -1,0 | -1,0 | | | |
| | 65-69 | -14,3 | -23,1 | -9,7 | 6,9 | 2,8 | -4,4 | 14,7 | -9,9 | 13,8 | -5,0 | | | |
| | 70 + | 1,9 | 6,2 | -10,1 | -12,1 | -8,2 | 1,6 | -2,4 | -1,0 | 3,5 | -4,3 | | | |
| | **Total** | **-13,7** | **-6,0** | **-7,2** | **-7,0** | **-2,8** | **-3,2** | **-6,5** | **-6,2** | **-5,3** | **-14,3** | **-5,7** | **-17,9** | **-17,4** |
| **Total** | | **-14,9** | **-9,3** | **-9,1** | **-10,9** | **-5,2** | **-4,7** | **-8,9** | **-9,1** | **-8,4** | **-14,9** | **-8,6** | **-21,3** | **-18,7** |

On observe un sous-dénombrement net pour chacune des provinces et territoires, les taux variant de 4,7 % à 21,3 %. L'ampleur du sous-dénombrement est plus importante aux Territoires du Nord-Ouest ainsi qu'au Nunavut. On doit mentionner qu'un faible taux de sous-dénombrement global n'est toutefois pas garant d'une meilleure couverture. Par exemple, le taux de glissement général observé à l'Île-du-Prince-Édouard (-9,3 %) cache le pire cas de surdénombrement à l'échelle provinciale des groupes d'âge-sexe (39,3 % chez les filles de 0 à 6 ans) et les troisième et quatrième pires cas de sous-dénombrement (respectivement de 34,8 % et 31,6% chez les femmes et les hommes de 18 à 24 ans). En revanche, bien que le taux de glissement général observé en Colombie-Britannique soit de -14,9%, cette province n'affiche que le treizième pire cas de sous-dénombrement (25,6% chez les hommes de 25 à 34 ans). On remarque également que le schéma de variation des taux de glissement pour les groupes d'âge-sexe diffère sensiblement d'une province à l'autre.

**Tableau 3.3**
**Taux de glissement pour les provinces et les territoires selon la taille de ménage**

| Province ou territoire | Taux de glissement | | | |
|---|---|---|---|---|
| | Ménages | Ménages d'une personne | Ménages de deux personnes | Ménages de trois personnes et plus |
| **Canada** | **-6,0** | **-4,0** | **-6,1** | **-7,3** |
| Terre-Neuve-et-Labrador | -10,1 | 0,8 | -0,3 | -20,3 |
| Île-du-Prince-Édouard | -6,3 | -4,6 | 6,2 | -16,2 |
| Nouvelle-Écosse | -7,4 | -2,8 | -5,8 | -11,4 |
| Nouveau-Brunswick | -6,7 | 0,2 | 1,0 | -16,4 |
| Québec | -5,3 | -10,9 | 0,0 | -5,4 |
| Ontario | -4,9 | 0,0 | -12,9 | -2,1 |
| Manitoba | -4,4 | 4,8 | -6,0 | -9,6 |
| Saskatchewan | -7,0 | -4,5 | -2,0 | -13,1 |
| Alberta | -4,1 | -0,7 | 2,6 | -10,9 |
| Colombie-Britannique | -11,0 | -4,1 | -9,5 | -16,8 |
| Yukon | -12,2 | -32,1 | -3,9 | -4,7 |
| Territoires du Nord-Ouest | -21,6 | -17,9 | -5,0 | -29,6 |
| Nunavut | -17,7[7] | .. | .. | .. |

On observe une sous-estimation du nombre de ménages de 6 % à l'échelle nationale. Ce taux de glissement est comparable au taux de glissement pour ce qui est des personnes qui est de -7,3 %.

À l'échelle provinciale et territoriale, il y a une variation importante du taux de glissement pour les ménages d'une personne. Le taux varie de -32,1 % pour le Yukon jusqu'à 4,8 % pour le Manitoba. Le même phénomène est observable pour les ménages de deux personnes et de trois personnes et plus mais avec une variation moindre. Pour les ménages de deux personnes, les taux de glissement varient de -12,9 % pour l'Ontario à 6,2 % pour l'Île-du-Prince-Édouard. Pour les ménages de trois personnes et plus, les taux de glissement varient de -29,6 % pour les Territoires du Nord-Ouest à -2,1% pour l'Ontario. Notons que l'on observe toujours un sous-dénombrement pour la catégorie des ménages de trois personnes et plus. À l'exception du Québec, de l'Ontario et du Yukon, le sous-dénombrement le plus élevé est observé pour les ménages de trois personnes et plus.

## 3.2 L'ajustement pour l'erreur de couverture de la population et du nombre de ménages

Pour corriger le problème de couverture illustré dans les tableaux 3.1 et 3.2 et réduire le biais qui en découle, on ajuste les données de l'enquête lors de la pondération en utilisant les estimations démographiques pour les groupes d'âge-sexe définis dans ces tableaux, et ce pour chaque province. Pour plus de détails sur la méthodologie de l'ajustement, voir la référence [1]. Cet ajustement réduit considérablement le biais engendré par les erreurs de couverture mais il n'élimine pas complètement le biais si les

[7] Pour le Nunavut, seul le nombre total de ménages est utilisé.

caractéristiques des individus que l'on a omis de l'enquête diffèrent de celles des individus inclus pour un même groupe d'âge dans une province.

On notera également que l'efficacité de l'ajustement de couverture à partir des estimations démographiques dépend beaucoup de la qualité de ces estimations démographiques et de leur exactitude à bien représenter la population cible de l'enquête. Les estimations démographiques ne sont pas exemptes d'erreurs. Ce sont des estimations post-censitaires basées sur les comptes de population du recensement de 1996, ajustées pour le sous-dénombrement net et qui tiennent compte des récentes statistiques sur la migration, les naissances, la mortalité, etc. Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institutions. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDM en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Pour corriger le problème de représentativité de l'échantillon par rapport au nombre de ménages selon leur taille, illustré dans le tableau 3.3, on ajuste les données de l'enquête à l'aide de données auxiliaires. En ajustant les poids de l'EDM pour refléter les estimations post-censitaires du nombre de ménages selon la taille, on vise à compenser le biais qui découle d'une représentation inadéquate des ménages. Toutefois on n'éliminera pas nécessairement ce biais si les caractéristiques des ménages non interviewés (omis ou non-répondants) diffèrent de celles des ménages répondants pour une même taille de ménage. Comme dans le cas des estimations démographiques de population, l'efficacité de l'ajustement dépend de la qualité des données auxiliaires sur le nombre de ménages. À ce titre, on notera que certains changements ont été apportés à la méthodologie utilisée pour la production des données auxiliaires sur le nombre de ménages pour les Territoires du Nord-Ouest et le Nunavut. L'impact de ces changements sur les résultats de tests statistiques visant la comparaison des estimations de l'EDM 2001 avec celles de l'EDM 1999 pour ces deux territoires n'a toutefois pas été étudié.

En plus des estimations démographiques quant aux groupes âge-sexe par province, trois autres ensembles de données auxiliaires sont utilisés pour ajuster les données d'enquêtes lors de la pondération en vue d'en améliorer la représentativité. Le premier ensemble de données est utilisé pour contrôler le nombre d'enfants et d'adultes dans certaines grandes villes. Le second est conçu pour contrôler le nombre de ménages monoparentaux et de couples avec enfants par province. Finalement, des comptes par grandes catégories de revenu en salaires et traitements sont utilisés lors de l'ajustement des poids afin d'assurer une certaine cohérence entre les distributions de revenu de l'EDM et celles provenant de sources externes.

## 4. Les erreurs de réponse

Les erreurs de réponse correspondent au manque d'exactitude des réponses aux questions. Elles peuvent être attribuables à différents facteurs parmi lesquels se retrouvent un questionnaire qui nécessite des améliorations, une mauvaise interprétation des questions par les intervieweurs ou les répondants ainsi que des erreurs dans les déclarations des répondants.

Dans l'EDM, les erreurs dans les déclarations des répondants peuvent survenir pour différentes raisons. Premièrement, il y a les erreurs de rappel qui se produisent lorsque le répondant oublie les dépenses effectuées durant la période de l'enquête (qui correspond à l'année civile) ou fournit une valeur erronée à cause de l'intervalle de temps écoulé entre le moment de l'achat et la date de l'entrevue. Les erreurs de rappel sont probablement la source d'erreur de réponse la plus importante de l'enquête étant donné que la période de référence est longue (12 mois) et qu'une grande variété d'information est demandée.

Une des principales mesures utilisées pour minimiser l'erreur de rappel dans l'EDM est de calculer, pour chaque ménage, la différence entre les entrées d'argent (revenus et autres montants reçus par le ménage) et les déboursés (dépenses plus variation nette de l'actif et du passif). Lorsque la différence excède 10 % des entrées d'argent ou des déboursés, le montant le plus élevé étant retenu, les répondants sont contactés à nouveau pour obtenir des renseignements supplémentaires et tenter d'identifier les erreurs ou les omissions. On encourage également le répondant à consulter divers documents (factures, relevés bancaires,...) pour fournir des données plus exactes. Pour déterminer les dépenses pour de petits articles achetés à intervalles réguliers, les intervieweurs suggèrent généralement aux répondants d'estimer la fréquence des achats et le prix généralement payé pour dériver les dépenses sur une période de 12 mois.

Une seconde source d'erreur dans la déclaration des répondants est l'erreur télescopique qui consiste à inclure dans la période de référence des événements qui se sont produits avant ou après celle-ci. Dans l'EDM, on considère que le fait d'utiliser l'année civile est un bon repère pour le début de la période de référence. De plus, comme la période de référence est longue, l'impact de l'erreur télescopique est moins important.

Les réponses par procuration peuvent également contribuer à l'erreur de réponse. Le membre du ménage qui a effectué une dépense est généralement la personne qui peut la rapporter avec la meilleure précision. C'est sûrement le cas par exemple pour les achats personnels. Les dépenses rapportées par personne interposée sont plus susceptibles d'être entachées d'une erreur de réponse et ce type d'erreur risque d'affecter plus sérieusement certaines catégories de dépenses.

Parmi d'autres sources d'erreur de réponse, il ne faut pas oublier le degré de coopération du répondant. Il se peut que le répondant décide pour des raisons personnelles de ne pas mentionner certaines dépenses ou de déformer les faits.

Dans l'EDM, on considère également que le fardeau imposé au répondant, étant donné la longueur de l'entrevue et la grande variété d'articles à rapporter, ainsi que le rythme

de l'entrevue peuvent entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Le temps d'entrevue varie beaucoup d'un ménage à l'autre selon la taille du ménage, le revenu et plusieurs autres caractéristiques. Pour certains ménages, l'entrevue peut prendre plus de cinq heures.

Il convient de signaler que l'on a ajouté des questions à l'EDM de 2001. Entre autres, pour 2001 seulement, des questions supplémentaires à l'EDM ont été ajoutées pour que les données de l'enquête puissent servir à la pondération de l'Indice des prix à la consommation. Ce changement peut nuire aux comparaisons historiques à l'égard de certaines variables. Par exemple, on a ajouté des questions à la rubrique « Produits de soins personnels » afin de recueillir des renseignements précis sur les produits de soins, le maquillage, les parfums, les désodorisants et les produits pour l'hygiène buccale. Ainsi, il se peut que les répondants aient donné un nombre accru de renseignements précis et que l'estimation à la hausse pour les « Produits de soins personnels » soit attribuable, du moins en partie, à la capacité accrue qu'ont les répondants de se souvenir de certains détails. L'effet des questions supplémentaires sur les estimations est difficile à quantifier.

Quoique l'on considère que les erreurs de réponses sont une source importante d'erreur dans une entrevue rétrospective, elles demeurent l'aspect lié à la qualité des données le plus difficile à mesurer. Il faut généralement entreprendre des études spéciales très coûteuses pour tenter de les mesurer. On tente de les réduire en implantant des techniques de sondage développées à cet effet.

## 5. Les erreurs de traitement

Des erreurs peuvent se produire lors de toutes les manipulations des données. Les principales étapes du traitement des données sont le codage des réponses, la saisie des données, la vérification, l'imputation des non-réponses partielles et la pondération. Dans l'EDM, différentes procédures sont appliquées à chacune des étapes afin de réduire au minimum les erreurs de traitement et les estimations de l'enquête sont comparées avec d'autres sources avant d'être publiées. Les erreurs qui peuvent découler des ajustements effectués lors de la pondération ont été décrites dans les sections 2 et 3. La présente section couvre principalement les autres types d'erreurs de traitement.

Le codage des réponses est nécessaire pour très peu de questions. Ce travail est effectué par l'intervieweur, puis vérifié par l'intervieweur principal. Avant 2001, la saisie des données était vérifiée à l'aide d'un système de vérification automatisé qui consistait à grouper les questionnaires par lots et à choisir, dans chacun de ces lots, un échantillon de questionnaires qui seraient saisis une seconde fois. Toutes les erreurs relevées étaient corrigées. Si le nombre d'erreurs dépassait un certain seuil, le lot complet était de nouveau soumis à la saisie. Contrairement aux années précédentes aucune procédure de vérification des questionnaires par lot n'a été utilisée pour 2001, suite à l'introduction d'un nouveau système de saisie des données (BLAISE). Toutefois, des règles de vérifications ont été implantées dans ce système afin de pouvoir assurer la cohérence des informations saisies. Les résultats préliminaires d'une étude sur la saisie semblent démontrer que les taux d'erreurs de saisie avec le nouveau système sont similaires à ceux avec l'ancien système.

Une première étape de vérification automatisée des questionnaires est effectuée après que chacun d'eux ait été vérifié manuellement par l'intervieweur et l'intervieweur principal. On s'assure que certaines règles essentielles de cohérence entre les réponses rapportées sur le questionnaire sont respectées. On identifie également les situations inhabituelles qui pourraient justifier des corrections. Cette étape de vérification automatisée est effectuée dans les bureaux régionaux de Statistique Canada, ce qui permet de communiquer avec les répondants lorsque des renseignements supplémentaires sont nécessaires pour résoudre des incohérences dans les réponses qu'ils ont fournies. Les problèmes identifiés au cours de cette vérification sont résolus par les membres des équipes de résolution des vérifications qui ont reçu une formation spéciale à cet égard. Par la suite, d'autres vérifications des données sont faites au bureau central et les réponses invalides sont corrigées.

Le traitement des données de l'EDM comprend également l'imputation de données pour la non-réponse partielle. Il y a non-réponse partielle lorsque le répondant refuse de répondre ou ne connaît pas la réponse à certaines questions particulières. L'approche pour imputer les données diffère selon qu'il s'agit de données catégoriques ou continues. Les données catégoriques peuvent prendre uniquement quelques valeurs (comme les questions pour lesquelles la réponse est oui ou non et les questions sur le type de logement habité) alors que les données continues peuvent prendre n'importe quelle valeur numérique (comme les revenus et les dépenses).

Les données catégoriques qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées à l'aide d'une technique "hot deck" où un ménage donneur est choisi de façon aléatoire parmi un groupe de ménages répondants possédant des caractéristiques semblables.

Les données sur le revenu et les dépenses sont quant à elles imputées au moyen de la technique du plus proche voisin. L'imputation se fait sur un groupe de variables à la fois, ce groupe étant formé de sorte à tenir compte des relations entre ces variables. Un groupe correspond généralement à une section du questionnaire. Pour chaque groupe, les valeurs manquantes d'un receveur (un ménage qui a des données manquantes pour au moins une de ces variables) sont imputées à partir des données du plus semblable parmi tous les donneurs (les ménages qui n'ont aucune donnée manquante pour ces variables). Pour chaque receveur, on détermine le donneur le plus semblable comme étant celui qui a la valeur minimale pour une certaine fonction de distance. Cette fonction est basée sur des variables d'appariement choisies parce qu'elles sont corrélées avec les variables à imputer. Par exemple, le revenu total du ménage fait partie des variables d'appariement pour presque toutes les sections portant sur les dépenses. Dans le choix du donneur, on tient compte également du fait que le receveur doit satisfaire certaines règles de cohérence après l'imputation de ces données manquantes. Pour la plupart des sections, l'imputation est effectuée au niveau du ménage mais dans certains cas, par exemple pour le revenu et les dépenses en vêtements, l'imputation est effectuée au niveau de la personne car c'est à ce niveau que l'on collecte l'information pour ces variables.

Notons qu'en 2001, l'imputation des données sur le revenu et les dépenses a été effectuée avec un nouveau système utilisant une méthodologie quelque peu différente de celle du système utilisé auparavant. Le nouveau système permet une meilleure utilisation des données catégoriques lors de la sélection du donneur. Des tests ont été

effectués avant l'introduction du système et les résultats obtenus étaient semblables à ceux obtenus avec l'ancien système. De plus, les résultats de cette année se comparent à ceux des années précédentes.

Le biais des estimations causé par l'imputation de la non-réponse partielle est difficile à évaluer. Il dépend de la différence entre les répondants et les non-répondants ainsi que de la capacité de la méthode d'imputation à produire des estimations non biaisées. Toutefois, les taux d'imputation donnent une indication de l'importance des non-réponses partielles. Ils sont présentés dans la section suivante.[8]

## 5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale

Une première indication de l'ampleur de la non-réponse partielle est la proportion de ménages nécessitant imputation et le nombre de variables imputées par ménage. Le questionnaire peut se diviser en deux grands groupes de variables, celles qui sont recueillies au niveau du ménage et celles collectées auprès des individus telles que le revenu et les dépenses en vêtements. Pour ces dernières catégories, il est important de noter que l'on accepte que le répondant fournisse seulement le revenu total ou les dépenses totales s'il est incapable de fournir la ventilation par source de revenu ou type de dépenses. Le niveau d'imputation pour les composantes du revenu et des dépenses vestimentaires est donc plus élevé mais n'affecte pas le revenu total, les dépenses vstimentaires totales ainsi que le total des dépenses.

Le pourcentage de ménages pour lesquels il a fallu imputer des dépenses de ménage (excluant les dépenses vestimentaires) est présenté dans la prochaine sous-section. Dans celle qui la suit, on présente le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable reliée aux dépenses vestimentaires et le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée. À la suite de l'imputation des données par le système, certaines corrections peuvent s'avérer nécessaires, autant sur les variables imputées par le système que sur les autres, afin d'assurer la cohérence des données. Dans les faits, cette année, environ 1 % des ménages et 2 % des personnes ont été modifiés manuellement. Les résultats sont fournis tant à l'échelle nationale que provinciale et territoriale. Cela fournit une indication des provinces et territoires qui sont les plus affectées par l'imputation.

### *5.1.1 Imputation de dépenses du ménage par province et territoire*

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable de dépenses (excluant les dépenses vestimentaires) figure dans le tableau 5.1. Les ménages utilisables correspondent à tous les ménages demeurant dans des logements admissibles, à l'exception de ceux qui n'ont pu être contactés, qui ont refusé de participer à l'enquête, qui ont fourni des données incomplètes ou qui ne balançaient pas (voir définitions dans la section 2.1). Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 251) pour un ménage.

[8] Pour des raisons opérationnelles, ces indicateurs de qualité ne sont pas disponibles pour les données catégoriques comme les caractéristiques du logement et l'équipement ménager.

**Tableau 5.1**
**Ménages qui nécessitent l'imputation de dépenses par province et territoire**

| Province | Ménages ( %) qui nécessitent l'imputation de **VARIABLES DE DÉPENSES** (sauf pour les dépenses en vêtements) | | | |
|---|---|---|---|---|
| | **Nombre de variables imputées (parmi 251)** | | | **TOTAL** |
| | **1** | **2** | **3 ou plus** | |
| **Canada** | **5,4** | **1,0** | **0,8** | **7,2** |
| Terre-Neuve-et-Labrador | 9,9 | 0,8 | 0,2 | 11,0 |
| Île-du-Prince-Édouard | 2,9 | 0,5 | 0,8 | 4,1 |
| Nouvelle-Écosse | 7,0 | 1,2 | 1,2 | 9,4 |
| Nouveau-Brunswick | 3,8 | 0,6 | 0,5 | 4,9 |
| Québec | 2,5 | 0,6 | 0,4 | 3,5 |
| Ontario | 7,2 | 2,0 | 2,0 | 11,3 |
| Manitoba | 4,3 | 0,7 | 0,6 | 5,6 |
| Saskatchewan | 6,9 | 1,0 | 0,8 | 8,7 |
| Alberta | 3,6 | 0,7 | 0,6 | 4,9 |
| Colombie-Britannique | 4,4 | 1,2 | 0,8 | 6,4 |
| Yukon | 6,0 | 0,4 | 0,4 | 6,8 |
| Territoires du Nord-Ouest | 8,9 | 2,8 | 0,3 | 12,0 |
| Nunavut | 9,1 | 1,0 | 0,0 | 10,0 |

Le tableau 5.1 montre qu'il a fallu imputer des dépenses (mise à part la section sur les vêtements) pour 7,2 % des ménages à l'échelle nationale, mais pour près de 75 % d'entre eux, seulement une variable a été imputée. Il y a très peu de ménages à l'échelle nationale qui ont eu plus d'une variable imputée (1,8 %). Les provinces pour lesquelles on observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable de dépense sont le Québec (3,5 %), l'Île-du-Prince-Édouard (4,1 %) et le Nouveau-Brunswick (4,9 %). On retrouve les taux les plus élevés à Terre-Neuve-et-Labrador (11 %), en Ontario (11,3 %) et aux Territoires du Nord-Ouest (12 %). Le faible pourcentage de ménages pour lesquels des variables doivent être imputées, combiné à un nombre généralement peu élevé de variables à imputer lorsque l'imputation est nécessaire, suggère que l'impact des valeurs imputées sur les estimations ne devrait pas être trop élevé.

*5.1.2 Imputation de dépenses vestimentaires et de revenu par province et territoire*

Étant donné qu'une partie des répondants fournissent uniquement le total pour les variables de dépenses en vêtements et les variables de revenu, celles-ci sont imputées, au niveau des individus, en deux étapes. Les individus pour lesquels il faut imputer seulement quelques composantes sont imputés en premier lieu, suivis par ceux pour lesquels seulement les totaux sont disponibles et qui nécessitent l'imputation de toutes les composantes. (Voir la référence [1] pour une description plus détaillée de ce processus).

Le pourcentage d'individus utilisables (personnes membres des ménages utilisables) pour lesquels il a fallu imputer au moins une variable de revenu est indiqué par province et territoire dans le tableau 5.2. On y retrouve le pourcentage de personnes pour lesquelles exactement une variable a été imputée, ceux dont deux variables ou plus (mais pas toutes) ont été imputées et le pourcentage des personnes pour lesquelles seul le revenu total était disponible, nécessitant ainsi l'imputation de toutes les

composantes. Le pourcentage total de personnes pour lesquelles une forme d'imputation de revenu a été effectuée est également donné. La dernière colonne du tableau 5.2 contient ce même pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée, mais pour les dépenses en vêtements.

**Tableau 5.2**
**Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province et territoire**

| Province | Pourcentage de personnes qui nécessitent l'imputation de VARIABLES DE REVENU | | | | Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 VARIABLES DE DÉPENSES VESTIMENTAIRES |
|---|---|---|---|---|---|
| | 1 variable imputée | 2 variables imputées ou plus (pas toutes) | Toutes les variables imputées (revenu total connu) | **TOTAL** (toute forme d'imputation) | |
| **Canada** | **0,5** | **0,2** | **2,7** | **3,5** | **12,6** |
| Terre-Neuve-et-Labrador | 0,2 | 0,1 | 1,7 | 2,0 | 14,8 |
| Î.-P.-É. | 0,0 | 0,1 | 5,1 | 5,2 | 13,3 |
| N.-É. | 0,5 | 0,3 | 3,1 | 3,9 | 15,2 |
| N.-B. | 0,2 | 0,2 | 5,9 | 6,4 | 13,1 |
| Québec | 0,3 | 0,1 | 1,4 | 1,8 | 9,8 |
| Ontario | 1,3 | 0,5 | 2,4 | 4,2 | 13,5 |
| Manitoba | 1,2 | 0,2 | 2,3 | 3,7 | 13,2 |
| Saskatchewan | 0,4 | 0,2 | 2,0 | 2,5 | 11,9 |
| Alberta | 0,2 | 0,3 | 3,2 | 3,6 | 15,1 |
| C.-B. | 0,6 | 0,3 | 2,9 | 3,8 | 9,8 |
| Yukon | 0,3 | 0,0 | 1,2 | 1,5 | 12,8 |
| T.N.-O. | 0,0 | 0,0 | 1,8 | 1,8 | 8,5 |
| Nunavut | 0,1 | 0,0 | 0,1 | 0,2 | 7,6 |

À partir de ces résultats, on constate qu'il y a moins de 4 % des personnes des ménages utilisables pour lesquelles il a fallu imputer au moins une variable de revenu. Pour environ 80 % d'entre elles, le revenu total était disponible mais toutes les composantes ont dû être imputées. Pour une très grande partie des autres personnes qui ont nécessité imputation, seulement une composante de revenu (une variable) a été imputée. À l'échelle provinciale et territoriale, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu est également peu élevé, variant entre un minimum de 0,2 % pour le Nunavut et un maximum de 6,4 % pour le Nouveau-Brunswick.

Dans la dernière colonne du tableau, on constate qu'environ 13 % des personnes ont nécessité une imputation pour au moins une des variables de dépenses vestimentaires. Les taux provinciaux et territoriaux s'échelonnent de 7,6 % pour le Nunavut à environ 15 % (Terre-Neuve-et-Labrador, Nouvelle-Écosse et Alberta). Presque toutes ces personnes ont fourni le total de leurs dépenses vestimentaires, mais il a fallu imputer les composantes. Le niveau élevé d'imputation effectuée sur les composantes des dépenses vestimentaires implique que les estimations de ces composantes pourraient être grandement affectées par l'imputation, alors que l'effet sur les estimations du total sera négligeable.

# Références


[1] Tremblay J. et Arsenault, S. 2001. *La méthodologie de l'Enquête sur les dépenses des ménages.* Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.

[2] Wolter, K.M. 1985. *Introduction to Variance Estimation.* New York. Springer-Verlag.

[3] Statistique Canada, Division de la statistique de revenue. 2001. *Guide de l'utilisateur — Enquête sur les dépenses des ménages.* N° 62F0026MIF2002002 au catalogue. Ottawa.

[4] Statistique Canada, Division de la statistique de revenue. 2001. *Les habitudes de dépenses au Canada.* N° 62-202-XPF au catalogue. Ottawa.

[5] Statistique Canada, Division des méthodes d'enquêtes auprès des ménages. 2001. *Enquête sur la population active, Rapport sur les opérations.* Enquête 200112. Ottawa.

# Annexe A

# Notation algébrique

### 1. Ajustement pour la non-réponse

Le sous-poids (c-à-d. le poids de sondage ajusté pour la non-réponse) d'un ménage k, noté $w_k^{NR}$, est

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{taux_g} \quad avec \quad taux_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

où

$s_{g,r}$ est l'ensemble des ménages répondants dans le groupe de non-réponse g,

$s_{g,nr}$ est l'ensemble des ménages non-répondants (refus, non-contacts, données inutilisables) dans le groupe de non-réponse g, et

• $\pi_k^{-1}$ est le poids de sondage attribué au ménage k.

### 2. Calcul du taux de glissement

Le taux de glissement pour un groupe de contrôle c, noté $taux_c$, est,

$$taux_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR}\right) - t_c}{t_c}$$

où

$s_{c,r}$ est l'ensemble des ménages répondants dans le groupe de contrôle c,

$w_k^{NR}$ est le sous-poids du ménage k, et

$t_c$ est le total de la donnée auxiliaire pour le groupe de contrôle c.